Draft Comparative Effectiveness Review

Number XX

The Effectiveness of Disease-modifying Anti-rheumatic drugs (DMARDs) in Children with Juvenile Idiopathic Arthritis (JIA)

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

This information is distributed solely for the purposes of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the Agency for Healthcare Research and Quality. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services determination or policy.

Contract No.

Prepared by:

Investigators

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

This report is based on research conducted by the XXX (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. XXX). The findings and conclusions in this document are those of the author(s), who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the State Children's Health Insurance Program (SCHIP).

AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting Comparative Effectiveness Reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see http://effectivehealthcare.ahrq.gov/reference/purpose.cfm

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence.

Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input. Comparative Effectiveness Reviews will be updated regularly.

Acknowledgments

Technical Expert Panel

AHRQ Contacts

Contents

Executive Summary 8
Introduction 17
Background 17
Scope and Key Questions 18
Methods 23
Topic Refinement 23
Search Strategy 23
Study Selection 23
Analytic Framework 26
Data Extraction 27
Quality Assessment 27
Rating the Body of Evidence 28
Assessing Applicability 29
Data Synthesis 29
Peer Review Process 30
Results 31
Literature Search and Screening 31
Key Question 1. In children with juvenile idiopathic arthritis (JIA), does treatment with disease-modifying anti-rheumatic drugs (DMARDs), compared to conventional treatment (i.e., non-steroidal anti-inflammatory drugs [NSAIDs] or corticosteroids) with or without methotrexate, improve laboratory measures of inflammation or radiological progression, symptoms (e.g., pain, symptom scores), or health status (e.g., functional ability, mortality)? 34
Key Question 2. In children with JIA, what are the comparative effects of DMARDs on laboratory markers of inflammation or radiological progression, symptoms (e.g., pain, symptom scores), or health status (e.g., functional ability, mortality)? 44
Key Question 3. In children with JIA, does the rate and type of adverse events differ between the various DMARDs or between DMARDs and conventional treatment with or without methotrexate? 47
Key Question 4. How do the efficacy, effectiveness, safety, and adverse effects of treatment with DMARDs differ among the various subtypes of JIA? 55
Key Question 5. What is the validity, reliability, responsiveness, and feasibility of the clinical outcomes measures for childhood JIA that are commonly used in clinical trials or within the clinical practice setting? 55

Summary and Discussion 71
Future Research 80
Efficacy of DMARDs 80
Safety of DMARDs 81
Impact of DMARDs on Health Outcomes 82
References 83
Abbreviations 87

Tables

Table 1. DMARDs evaluated 21
Table 2. Conventional treatments evaluated 22
Table 3. Sources of citations 31
Table 4. Studies comparing biologic DMARDs versus conventional treatments with or without methotrexate 36
Table 5. Studies comparing non-biologic DMARDs versus conventional treatments with or without methotrexate 38
Table 6. Studies comparing various DMARDs with one another 45
Table 7 – Part 1. Adverse events reported in RCTs – dropouts and adverse events related to organ systems 49
Table 7 – Part 2. Adverse events reported in RCTs – specific symptoms 50
Table 7 – Part 3. Adverse events reported in RCTs – other 51
Table 8. Outcomes measures assessed 59
Table 9. Studies of psychometric properties of common JIA outcomes measures and developing definitions of treatment response 62
Table 10. Validity – correlations of instruments with measures of diseases and other instruments 66
Table 11. Responsiveness 68
Table 12. Summary of the evidence on comparative effectiveness and harms of DMARDs for childhood JIA 71
Table 13. GRADE summary table for Key Question 1 – biologic DMARDs 74
Table 14. GRADE summary table for Key Question 1 – non-biologic DMARDs 75
Table 15. GRADE summary table for Key Question 2 76
Table 16. GRADE summary table for Key Question 4 77
Table 17. Evidence summary table for Key Question 5 78

Figures

Figure 1. Analytic framework 26
Figure 2. Literature flow diagram 32
Figure 3. Treatment comparisons evaluated in efficacy studies 33
Figure 4. Comparison of symptomatic flare in children with JIA randomized to continuing a biologic DMARD versus placebo 42

Appendices

Appendix A. Exact Search Strings
Appendix B. Screening Criteria
Appendix C. Data Abstraction Forms
Appendix D. Evidence Tables – Main Literature Search
Appendix E. Adverse Events – Main Search and Horizon Scan
Appendix F. Excluded Studies

The Effectiveness of Disease-modifying Anti-rheumatic drugs (DMARDs) in Children with Juvenile Idiopathic Arthritis (JIA)

Executive Summary

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

Background

Juvenile idiopathic arthritis (JIA) is the most common rheumatologic disease in childhood, with an overall prevalence of 7 to 400 per 100,000 children. JIA is an important cause of chronic disease in childhood, with prevalence similar to type I diabetes mellitus. Several classification systems have been used over time to categorize the various subtypes of juvenile arthritis based upon clinical presentation and disease course. These classification schemes are complicated by the heterogeneity of disease and overlapping clinical features. Currently, the standard nomenclature is JIA, which consists of six main subtypes. These subtypes are useful in examining potential differences in treatment response and prognosis. The main subtypes of JIA are:

- Systemic arthritis: Initial presentation includes fever, rash, and arthritis; one-quarter of children may have severe destructive disease.
- Oligoarthritis: Affects up to four joints within the first 6 months of illness; may be persistent (i.e., involving no more than four joints) or extended (i.e., involving more than four joints after the first 6 months of illness), and may be associated with uveitis.
- Polyarthritis: Affects five or more joints during the first 6 months of disease; further subclassified as rheumatoid factor-negative and rheumatoid factor-positive disease, the latter of which may be associated with more destructive joint disease.
- Enthesitis-related arthritis: May be associated with uveitis.
- Psoriatic arthritis: May be associated with uveitis.
- Undifferentiated: Arthritis lasting more than 6 weeks that does not meet the criteria for any of the above subtypes, or that meets the criteria for more than one subtype.

JIA can place a severe physical and psychological burden on affected children and be a major stressor to their families. As is true for all chronic conditions in childhood, treatment of JIA may

be enhanced through the use of a multidisciplinary team to address these issues. There is no cure for JIA, but over the past 25 years new therapies have provided great advances in treatment and symptom control. Previous treatments with non-steroidal anti-inflammatory drugs (NSAIDs; e.g., ibuprofen) and corticosteroids (systemic or intra-articular) were only partially effective in treating the symptoms of arthritis and reducing long-term complications (e.g., growth delay, erosive joint disease). Treatment with the class of agents known as disease-modifying anti-rheumatic drugs (DMARDs) has become an increasingly important component of care because these drugs appear to lead to better disease control, with higher numbers of children achieving remission, and fewer children suffering long-term joint damage. DMARDs interfere with the making or working of immune cells that cause joint inflammation and are typically classified as either biologic drugs, which are administered intravenously or subcutaneously, or non-biologic drugs, which are given orally or subcutaneously. In general, the non-biologic DMARDs are older. One of these non-biologic DMARDs, methotrexate, has been used for so long in the treatment of JIA that it is often considered part of conventional treatment, along with NSAIDs and intra-articular corticosteroids.

Although there is significant optimism that treatment with the newer biologic DMARDs may increasingly lead to long-term disease remission, there are many unanswered questions about the safety of these drugs, especially for long-term use in children. For example, the U.S. Food and Drug Administration (FDA) recently placed a black box warning on the entire class of biologic DMARDs targeting tumor necrosis factor (TNF) alpha, including etanercept, infliximab, and adalimumab, due to concerns about potential increased risk of malignancy, in particular lymphoma. There are also important questions about effectiveness, including the comparative effectiveness of DMARDs versus conventional treatment and the comparative effectiveness of the various DMARDs versus one another. Furthermore, it is possible that the effectiveness of these drugs varies by subtype of JIA. Understanding the circumstances in which a DMARD should be used, and which DMARD(s) should be selected, is challenging because JIA is heterogeneous across the various subtypes. A clear synthesis of the available evidence is needed to help clinicians provide care for children with JIA and to identify the important gaps in the scientific literature.

Juvenile arthritis has a broad impact on a child's physical and mental health. Developing instruments that accurately assess the effect of JIA on health and well-being is critical both to assessing the overall impact of the disease and to quantifying the efficacy of treatments. The heterogeneity of disease severity, the broad age range of affected individuals, and fluctuations in the natural history of the disease complicate measuring disease activity and treatment effects in children with JIA. To provide the most accurate assessment of treatment effects we depend on the performance characteristics of the outcomes measures reported in the scientific literature. Multiple instruments have been developed or adapted to assess severity of disease, disability, and quality of life in JIA. Understanding the reliability, validity, and responsiveness of these instruments will facilitate interpretation of clinical trial data.

This comparative effectiveness review summarizes the evidence on the benefits and harms of DMARDs compared to conventional treatment (NSAIDs and/or intra-articular corticosteroids) with or without methotrexate, and of the various DMARDs compared to one another, in children with JIA. In addition, this review summarizes the usefulness of selected tools commonly used to measure clinical outcomes associated with JIA.

Key questions addressed are:

Key Question 1. In children[a] with JIA,[b] does treatment with DMARDs,[c] compared to conventional treatment (i.e., NSAIDs or corticosteroids) with or without methotrexate,[d] improve laboratory measures of inflammation or radiological progression, symptoms (e.g., pain, symptom scores), or health status (e.g., functional ability, mortality)?

(a) "Children" are defined as individuals aged 18 years or younger.

(b) "JIA" includes any subtype of any severity of the following:
- JIA according to the International League of Associations for Rheumatology (ILAR) criteria;
- Juvenile rheumatoid arthritis (JRA) according to the American College of Rheumatology (ACR) definition; or
- Juvenile chronic arthritis (JCA) according to the European League Against Rheumatism (EULAR) criteria.

(c) DMARDs evaluated are: abatacept, adalimumab, anakinra, canakinumab, etanercept, infliximab, intravenous immunoglobulin (IVIG), rilonacept, rituximab, and tocilizumab (biologic DMARDs); and azathioprine, cyclosporine A, penicillamine, hydroxychloroquine, leflunomide, methotrexate, mycophenolate mofetil, sulfasalazine, tacrolimus (FK506), and thalidomide (non-biologic DMARDs).

(d) Conventional treatments evaluated are: betamethasone, triamcinolone acetonide, triamcinolone hexacetonide, celecoxib, etodolac, ibuprofen, indomethacin, meloxicam, naproxen, oxaprozin, and tolmetin.

Key Question 2. In children with JIA, what are the comparative effects of DMARDs[e] on laboratory markers of inflammation or radiological progression, symptoms (e.g., pain, symptom scores), or health status (e.g., functional ability, mortality)?

(e) This question is identical to Key Question 1, but focuses on comparisons of one DMARD versus another, rather than on comparisons of DMARDs versus conventional treatments.

Key Question 3. In children with JIA, does the rate and type of adverse events[f] differ between the various DMARDs or between DMARDs and conventional treatment with or without methotrexate?

(f) Because of the known risks associated with DMARDs, we focused primarily on serious infections and the development of cancer when assessing adverse events. Other adverse events considered included mortality, hepatitis, bone marrow suppression, nausea or vomiting, and risks to fetus or pregnant mother.

Key Question 4. How do the efficacy, effectiveness, safety, and adverse effects of treatment with DMARDs differ among the various subtypes[g] of JIA?

(g) Subtypes of JIA include:
- Systemic arthritis
- Oligoarthritis
- Polyarthritis
- Enthesitis-related arthritis
- Psoriatic arthritis
- Other (arthritis of unknown cause with symptoms lasting more than 6 weeks).

Key Question 5. What is the validity, reliability, responsiveness, and feasibility of the clinical outcomes measures[h] for childhood JIA that are commonly used in clinical trials or within the clinical practice setting?

(h) The outcomes measures assessed were those most commonly used in clinical trials and practice, as well as newer instruments of particular interest that were selected in consultation with the project's technical expert panel (TEP). The outcome measures assessed were:
- Measures of disease activity:
  - Active joint count (AJC)
  - Physician global assessment of disease activity (PGA)
  - Parent/patient global assessment of well-being (PGW)
- Measure of functional status/disability:
  - Childhood Health Assessment Questionnaire (CHAQ)
- Measures of health-related quality of life:
  - Child Health Questionnaire (CHQ)
  - Pediatric Quality of Life Inventory (PedsQL) 4.0
  - Pediatric Quality of Life Inventory Rheumatology Module (PedsQL-RM)
- Composite measures of response to therapy and developing definitions of disease status:
  - American College of Rheumatology Pediatric Response Criteria (ACR Pediatric 30)
  - A consensus-based definition of remission
  - Flare
  - Minimal disease activity (MDA)

These instruments were assessed for test-retest reliability, inter- and intra-rater reliability, internal reliability, construct validity, responsiveness (standardized response mean and responsiveness index), and feasibility metrics such as time to administer.

Conclusions

Table A provides an aggregated view of the strength of evidence and brief conclusions from this review of the comparative benefits and harms of DMARDs for children with JIA.

Table A. Summary of the evidence on comparative effectiveness and harms of DMARDs for childhood JIA

Key question	Strength of evidence	Conclusions
1. In children with JIA, does treatment with DMARDs, compared to conventional treatment:		
a. Improve laboratory measures of inflammation?	Low	Trials of DMARDs usually report changes in laboratory measures of inflammation (e.g., ESR). However, ESR is inconsistently associated with treatment. This is based on 14 studies of 1060 subjects.
b. Improve radiological progression?	Insufficient	Insufficient data are available to evaluate the impact of DMARDs on radiological progression. Only one cohort study of 63 subjects reported data on radiological progression.
c. Improve symptoms?	Moderate	Among children who have responded to a biologic DMARD, randomized discontinuation trials show that continued treatment for 4 months to 2 years decrease the risk of having a flare (RR 0.46, 95% CI 0.36 to 0.60). This is based on four studies of 322 subjects. Among the non-biologic DMARDs, there is some evidence that methotrexate is superior to conventional therapy and oral corticosteroids based on two randomized trials of 215 subjects.
d. Improve health status?	Low	Changes in health status were reported in 12 studies involving 927 subjects. Health status improved inconsistently with treatment with DMARDs.
2. In children with JIA, what are the comparative effects of DMARDs on:		
a. Laboratory measures of inflammation?	Low	Trials of DMARDs usually report changes in laboratory measures of inflammation (e.g., ESR). However, ESR is inconsistently associated with treatment. This is based on 4 RCTs of 448 subjects and 1 cohort study of 72 subjects.
b. Radiological progression?	Insufficient	No study addressed radiologic progression.
c. Symptoms?	Low	The non-biologic DMARDs that were compared directly (penicillamine vs. hydroxychloroquine, sulfasalazine vs. hydroxychloroquine, and leflunomide vs. methotrexate) had similar effectiveness. Changes in symptoms between the treatment arms were not measured with significant precision to detect a difference. This is based on 4 RCTs of 448 subjects and 1 cohort study of 72 subjects. One poor-quality RCT of 94 subjects found that etanercept was similar to infliximab.

Key question	Strength of evidence	Conclusions
d. Health status?	Low	The non-biologic DMARDs that were compared directly (penicillamine vs. hydroxychloroquine, sulfasalazine vs. hydroxychloroquine, and leflunomide vs. methotrexate) had similar effectiveness. Changes in health status between the treatment arms were not measured with significant precision to detect a difference. This is based on 4 RCTs of 448 subjects and 1 cohort study of 72 subjects. One poor quality RCT of 94 subjects found that etanercept was similar to infliximab.
3. In children with JIA, does the rate and type of adverse events differ between:		
a. The various DMARDs?	Insufficient	Three RCTs directly compared two DMARDs; two compared penicillamine to hydroxychloroquine, and one compared leflunomide to methotrexate. The rate and type of adverse events did not differ between treatment groups in these studies. High variability across studies in the ascertainment and reporting of adverse events preclude valid comparisons of the rate and type of adverse events among the various DMARDs.
b. DMARDs and conventional treatment with or without methotrexate?	Insufficient	No RCT directly compared a DMARD to conventional treatment. Thirteen trials directly compared a DMARD to placebo. The rate and type of adverse events were generally similar between intervention and placebo groups, with the notable exceptions of infliximab plus methotrexate being associated with more serious adverse events (32% vs. 5% over differing lengths of followup), and methotrexate being associated with higher rates of laboratory abnormalities (35% vs. 13%).
4. How do the efficacy, effectiveness, safety, and adverse effects of treatment with DMARDs differ among the various categories of JIA?	Insufficient	Only one study – an RCT of methotrexate versus placebo in which each group could also receive oral corticosteroids, intra-articular corticosteroids, and NSAIDs – evaluated efficacy by JIA subtype. No difference was found among those with extended oligoarticular JIA (n = 43) and systemic JIA (n = 45). We did not identify any studies that provide reliable information on the comparative safety or rates or types of adverse events among the various subtypes of JIA.
5. What is the validity, reliability, responsiveness, and feasibility of the clinical outcome measures for childhood JIA that are commonly used in clinical trials or within the clinical practice setting?	Insufficient	Most of the studies examining the psychometric properties of the instruments used in JIA were fair-quality cross-sectional or longitudinal non-randomized controlled trials. No one instrument or outcomes measure appeared superior in measuring disease activity or functional status. The current response criteria of the ACR Pediatric 30, a composite measure that includes articular indices, functional status, laboratory measure, and global

Key question	Strength of evidence	Conclusions
		assessments, takes into account the various measures most commonly used. However, the responsiveness of several of these measures, including functional status and parent/patient global assessment, are poor to moderate, and they may not adequately reflect changes in disease state. Furthermore, given that the ACR Pediatric 30 is a relative measure of disease activity, the impact of JIA subtype on percent improvement is unclear, as certain instruments, such as the CHAQ, appear to have differential responsiveness depending on extent of disease at baseline. The ACR Pediatric 30 is also a relative measure of disease activity and not a measure of current disease state.

Abbreviations: ACR = American College of Rheumatology; CHAQ = Childhood Health Assessment Questionnaire; CI = confidence interval; DMARD(s) = disease-modifying anti-rheumatic drug(s); ESR = erythrocyte sedimentation rate; JIA = juvenile idiopathic arthritis; NSAIDs = non-steroidal anti-inflammatory drugs; RCT = randomized controlled trial; RR = risk ratio

Remaining Issues

Despite the importance of DMARDs for the treatment of childhood JIA, there is a paucity of comparative evidence for long-term benefits and harms. One particular important challenge is the development of outcome measure tools that fully describe the impact of the condition and that are both feasible to administer and sensitive to changes in the status of the condition.

Future Research

Although DMARDs have improved health outcomes for children with JIA, few data are available to evaluate the comparative effectiveness of either specific DMARDs or general classes of DMARDs (e.g., non-biologic vs. biologic, or by mechanism of action). Research on the effectiveness of treatments for JIA is challenging because it is a rare condition that includes multiple subtypes, which could potentially respond differentially to therapy. Furthermore, the health impact of JIA fluctuates over time. Therefore, trials require large sample sizes with long follow-up periods.

Developing a summary estimate of effectiveness of the DMARDs is challenging because there is:

- Heterogeneity in the study population. Changes in the definition of JIA (e.g., JRA, JCA) may have led to the inclusion in studies of individuals who may respond differently to treatments. Similarly, differences by disease subtype (e.g., polyarticular, pauciarticular, systemic) might lead to different conclusions about the effectiveness of treatment.
- Variation in comparators. Over time, the standard of care for JIA has changed. For example, relatively recent studies of biologic DMARDs often allow methotrexate, a DMARD, in the comparator group, while older studies do not include methotrexate in the comparator groups. Some older studies included systemic corticosteroids as a comparator.

- Outcome measures vary across the studies and are sometimes incompletely described. Some studies report the percentage improvement from baseline without providing baseline data or an estimate of variability. Among six randomized discontinuation trials identified for this review, four reported laboratory measures of inflammation, four reported whether a flare occurred, three reported active joint count, and four reported quality of life as measured by CHAQ. Of those that reported the CHAQ score, one reported only the percentage change from baseline without the absolute value or measure of dispersion (e.g., range, standard deviation), and two gave average values without measures of dispersion.

Future trials in this domain should consider:

- The challenge of the appropriate comparator. Trials are needed to evaluate the effectiveness of DMARDs compared to conventional therapy as well as against other DMARDs. Defining conventional therapy is challenging because it evolves with advances in the field. Factorial designs involving multiple treatments are a potential solution. Patient-level meta-analysis, pre-planned across different trials, may also help address this issue.
- The issue of treatment-by-subtype interaction. To fully explore comparative effectiveness, larger studies will be needed. In addition, patient-level meta-analysis may help address this challenge.
- The need for study populations who are representative of typical patients with JIA. Subjects from the studies included in this review were identified through specialty clinics, which is appropriate for rare conditions. However, baseline characteristics varied. Studies should be designed to reflect the comparative effectiveness for typical subjects at various points along the disease spectrum (e.g., at presentation, after failing conventional treatment).
- The variable course of JIA. Trials that evaluate the efficacy of treatment should be sufficiently long with frequent assessment of health status to capture the natural variability of the disease course.
- The need for standardized definitions for, and systematic ascertainment and reporting of, adverse events possibly associated with therapeutic interventions in the treatment of JIA.
- The impact of DMARDs on the specific health conditions associated with JIA, including uveitis and macrocyte activation syndrome.

Study designs other than randomized controlled trials (RCTs) will be important in understanding the role of DMARDs in JIA. Randomized discontinuation trials have helped to define the risk of flare in patients who respond to a particular DMARD. Large cohort studies will be important for evaluating the risk of adverse events associated with DMARDs. Such studies could also be important for better characterizing long-term outcomes in JIA.

Few high-quality data are available regarding the adverse events associated with DMARDs. Because JIA is a chronic illness, understanding the long-term adverse effects of these drugs is critical. One solution to evaluating risk would be to develop registries for DMARDs when used for childhood JIA. Understanding such risk will also provide information about the sequence in which these drugs should be used for difficult-to-treat JIA, or the impact of using multiple drugs.

Our findings suggest that short-term mortality rates associated with DMARDs are very low (we identified only a single patient among more than 3000 who died shortly after receiving a DMARD). The incidence of malignancies during a short course of DMARD treatment also appears to be very low. However, the available evidence is inadequate to determine whether the rates and types of adverse events differ between the various DMARDs or between DMARDs and conventional treatment. The findings from RCTs do not reveal a clear pattern pertaining to adverse events associated with the treatment of JIA with DMARDs compared to placebo. A horizon scan of other study designs revealed marked differences in the rate and type of adverse event by DMARD, but these findings should be interpreted with caution for several reasons, including: variable definitions of adverse events across studies; non-systematic methods of ascertaining adverse events; nearly universal lack of standard reporting of serious adverse events; a predominance of case reports and uncontrolled series; small sample sizes in most series and RCTs; a limited number of studies for many individual DMARDs; and frequent use of multiple medications and other co-interventions.

Finally, our findings suggest the need for better clinical outcomes measures that are responsive to change across the full spectrum of disease severity. Consistent use of such outcomes measures would facilitate comparative effectiveness research.

The heterogeneity in disease severity and the broad impact of the disease on both physical and psychosocial aspects of children's lives make it difficult to accurately assess children using one instrument or measure. Given the complex nature of JIA, with the potential for both chronic and acute functional limitations and pain, it is difficult to find one tool or instrument that can be responsive to all the facets of disease. Efforts to develop a more standardized composite measure which could incorporate articular indices, severity, and a broader assessment of functional limitations and psychosocial impact would be useful to better differentiate levels of disease activity and overall impact of disease. The current response criteria of the ACR Pediatric 30 definition of improvement, a composite measure which includes articular indices, functional status, laboratory measure, and global assessments, takes into account the various measures most commonly used. However, the responsiveness of several of these measures, including functional status and parent/patient global assessment, are poor to moderate, and they may not adequately reflect changes in disease state. Furthermore, given that the ACR Pediatric 30 is a relative measure of disease activity, the impact of JIA subtype on percent improvement is unclear, as certain instruments, such as the CHAQ, appear to have differential responsiveness by extent of disease at baseline. The ACR Pediatric 30 is also a relative measure and not a measure of current disease state.

Developing an instrument or composite measure to accurately describe all the aspects of JIA, including disease activity, functional status, and quality of life would improve our understanding of the overall impact of JIA. In addition, focusing on the most responsive outcome measures to assess treatment effects would enhance our ability to detect promising new treatments.

Introduction

Background

Juvenile idiopathic arthritis (JIA) is the most common rheumatologic disease in childhood, with an overall prevalence of 7 to 400 per 100,000 children.[1,2] JIA is an important cause of chronic disease in childhood, with prevalence similar to type I diabetes mellitus.[3] Several classification systems have been used over time to categorize the various subtypes of juvenile arthritis based upon clinical presentation and disease course. These classification schemes are complicated by the heterogeneity of disease and overlapping clinical features. Currently, the standard nomenclature is JIA, which consists of six main subtypes. These subtypes are useful in examining potential differences in treatment response and prognosis. The main subtypes of JIA are:[4]

- Systemic arthritis: Initial presentation includes fever, rash, and arthritis; one-quarter of children may have severe destructive disease.
- Oligoarthritis: Affects up to four joints within the first 6 months of illness; may be persistent (i.e., involving no more than four joints) or extended (i.e., involving more than four joints after the first 6 months of illness), and may be associated with uveitis.
- Polyarthritis: Affects five or more joints during the first 6 months of disease; further subclassified as rheumatoid factor-negative and rheumatoid factor-positive disease, the latter of which may be associated with more destructive joint disease.
- Enthesitis-related arthritis: May be associated with uveitis.
- Psoriatic arthritis: May be associated with uveitis.
- Undifferentiated: Arthritis lasting more than 6 weeks that does not meet the criteria for any of the above subtypes, or that meets the criteria for more than one subtype.

JIA can place a severe physical and psychological burden on affected children and be a major stressor to their families. As is true for all chronic conditions in childhood, treatment of JIA may be enhanced through the use of a multidisciplinary team to address these issues. There is no cure for JIA, but over the past 25 years new therapies have provided great advances in treatment and symptom control. Previous treatments with non-steroidal anti-inflammatory drugs (NSAIDs; e.g., ibuprofen) and corticosteroids (systemic or intra-articular) were only partially effective in treating the symptoms of arthritis and reducing long-term complications (e.g., growth delay, erosive joint disease). Treatment with the class of agents known as disease-modifying anti-rheumatic drugs (DMARDs) has become an increasingly important component of care because these drugs appear to lead to better disease control, with higher numbers of children achieving remission, and fewer children suffering long-term joint damage. DMARDs interfere with the making or working of immune cells that cause joint inflammation and are typically classified as either biologic drugs, which are administered intravenously or subcutaneously, or non-biologic drugs, which are given orally or subcutaneously. In general, the non-biologic DMARDs are older. One of these non-biologic DMARDs, methotrexate, has been used for so long in the treatment of JIA that it is often considered part of conventional treatment, along with NSAIDs and intra-articular corticosteroids.

Although there is significant optimism that treatment with the newer biologic DMARDs may increasingly lead to long-term disease remission, there are many unanswered questions about the safety of these drugs, especially for long-term use in children. For example, the U.S. Food and Drug Administration (FDA) recently placed a black box warning on the entire class of biologic DMARDs targeting tumor necrosis factor (TNF) alpha, including etanercept, infliximab, and adalimumab, due to concerns about potential increased risk of malignancy, in particular lymphoma. There are also important questions about effectiveness, including the comparative effectiveness of DMARDs versus conventional treatment and the comparative effectiveness of the various DMARDs versus one another. Furthermore, it is possible that the effectiveness of these drugs varies by subtype of JIA. Understanding the circumstances in which a DMARD should be used, and which DMARD(s) should be selected, is challenging because JIA is heterogeneous across the various subtypes. A clear synthesis of the available evidence is needed to help clinicians provide care for children with JIA and to identify the important gaps in the scientific literature.

Juvenile arthritis has a broad impact on a child's physical and mental health. Developing instruments that accurately assess the effect of JIA on health and well-being is critical both to assessing the overall impact of the disease and to quantifying the efficacy of treatments. The heterogeneity of disease severity, the broad age range of affected individuals, and fluctuations in the natural history of the disease complicate measuring disease activity and treatment effects in children with JIA. To provide the most accurate assessment of treatment effects we depend on the performance characteristics of the outcomes measures reported in the scientific literature. Multiple instruments have been developed or adapted to assess severity of disease, disability, and quality of life in JIA. Understanding the reliability, validity, and responsiveness of these instruments will facilitate interpretation of clinical trial data.

In this comparative effectiveness review, we examine the scientific literature on DMARDs for JIA in childhood. Moreover, we review evidence regarding the usefulness of available outcomes measures for JIA that are commonly used in clinical trials and within the clinical practice setting.

Scope and Key Questions

This review summarizes the evidence on the benefits and harms of DMARDs compared to conventional treatment (NSAIDs and/or intra-articular corticosteroids) with or without methotrexate, and of the various DMARDs compared to one another, in children with JIA. In addition, this review summarizes the usefulness of selected tools commonly used to measure clinical outcomes associated with JIA.

Key questions addressed are:

Key Question 1. In children[a] with JIA,[b] does treatment with DMARDs,[c] compared to conventional treatment (i.e., NSAIDs or corticosteroids) with or without methotrexate,[d] improve laboratory measures of inflammation or radiological progression, symptoms (e.g., pain, symptom scores), or health status (e.g., functional ability, mortality)?

(a) "Children" are defined as individuals aged 18 years or younger.

(b) "JIA" includes any subtype of any severity of the following:
- JIA according to the International League of Associations for Rheumatology (ILAR) criteria;
- Juvenile rheumatoid arthritis (JRA) according to the American College of Rheumatology (ACR) definition; or
- Juvenile chronic arthritis (JCA) according to the European League Against Rheumatism (EULAR) criteria.

(c) DMARDs evaluated are: abatacept, adalimumab, anakinra, canakinumab, etanercept, infliximab, intravenous immunoglobulin (IVIG), rilonacept, rituximab, and tocilizumab (biologic DMARDs); and azathioprine, cyclosporine A, penicillamine, hydroxychloroquine, leflunomide, methotrexate, mycophenolate mofetil, sulfasalazine, tacrolimus (FK506), and thalidomide (non-biologic DMARDs).

(d) Conventional treatments evaluated are: betamethasone, triamcinolone acetonide, triamcinolone hexacetonide, celecoxib, etodolac, ibuprofen, indomethacin, meloxicam, naproxen, oxaprozin, and tolmetin.

Key Question 2. In children with JIA, what are the comparative effects of DMARDs[e] on laboratory markers of inflammation or radiological progression, symptoms (e.g., pain, symptom scores), or health status (e.g., functional ability, mortality)?

(e) This question is identical to Key Question 1, but focuses on comparisons of one DMARD versus another, rather than on comparisons of DMARDs versus conventional treatments.

Key Question 3. In children with JIA, does the rate and type of adverse events[f] differ between the various DMARDs or between DMARDs and conventional treatment with or without methotrexate?

(f) Because of the known risks associated with DMARDs, we focused primarily on serious infections and the development of cancer when assessing adverse events. Other adverse events considered included mortality, hepatitis, bone marrow suppression, nausea or vomiting, and risks to fetus or pregnant mother.

Key Question 4. How do the efficacy, effectiveness, safety, and adverse effects of treatment with DMARDs differ among the various subtypes[g] of JIA?

(g) Subtypes of JIA include:
- Systemic arthritis
- Oligoarthritis
- Polyarthritis
- Enthesitis-related arthritis
- Psoriatic arthritis
- Other (arthritis of unknown cause with symptoms lasting more than 6 weeks).

Key Question 5. What is the validity, reliability, responsiveness, and feasibility of the clinical outcomes measures[h] for childhood JIA that are commonly used in clinical trials or within the clinical practice setting?

(h) The outcomes measures assessed were those most commonly used in clinical trials and practice, as well as newer instruments of particular interest that were selected in consultation with the project's technical expert panel (TEP). The outcome measures assessed were:

- Measures of disease activity:
  - Active joint count (AJC)
  - Physician global assessment of disease activity (PGA)
  - Parent/patient global assessment of well-being (PGW)
- Measure of functional status/disability:
  - Childhood Health Assessment Questionnaire (CHAQ)
- Measures of health-related quality of life:
  - Child Health Questionnaire (CHQ)
  - Pediatric Quality of Life Inventory (PedsQL) 4.0
  - Pediatric Quality of Life Inventory Rheumatology Module (PedsQL-RM)
- Composite measures of response to therapy and developing definitions of disease status:
  - American College of Rheumatology Pediatric Response Criteria (ACR Pediatric 30)
  - A consensus-based definition of remission
  - Flare
  - Minimal disease activity (MDA)

These instruments were assessed for test-retest reliability, inter- and intra-rater reliability, internal reliability, construct validity, responsiveness (standardized response mean and responsiveness index), and feasibility metrics such as time to administer.

Table 1. DMARDs evaluated

Generic name	Biologic or non-biologic?	U.S. trade name(s)	Mechanism of action	FDA-approved for JIA?
Abatacept	Biologic	Orencia	Anti-CD28, T-cell costimulator antibodies	Yes
Adalimumab	Biologic	Humira	TNF inhibitor	Yes
Anakinra	Biologic	Kineret	IL-1 receptor antagonist	No
Canakinumab	Biologic	Ilaris	IL-1 blocker	No
Etanercept	Biologic	Enbrel	TNF inhibitor	Yes
Infliximab	Biologic	Remicade	TNF inhibitor	No
IVIG	Biologic	Baygam, Carimune NF, Flebogamma 5% DIF, Gammar P, Gamunex 10%, Gammagard S/D, Gammagard Liquid 10%, Gammar P, Iveegam EN, Octagam 5%, Panglobulin, Polygam S/D, Privigen 10%, Vivaglobin	Interaction with activating Fc receptors	No
Rilonacept	Biologic	Arcalyst	IL-1 blocker	No
Rituximab	Biologic	Rituxan	Binds to CD20 antigen	No
Tocilizumab	Biologic	Actemra	IL-6 receptor antagonist	No
Azathioprine	Non-biologic	Azasan; Imuran	Purine synthesis inhibitor	No
Cyclosporine A	Non-biologic	Neoral, Gengraf	Inhibits calcineurin	No
Penicillamine	Non-biologic	Depen; Cuprimine	Unknown (may lower IgM rheumatoid factor, depresses T-cell activity)	No
Hydroxy-chloroquine	Non-biologic	Plaquenil	Not well understood, may reduce T-lymphocyte transformation and chemotaxis	No
Leflunomide	Non-biologic	Arava	Isoxazole immunomodulatory agent	No
Methotrexate	Non-biologic	Methotrexate LPF	Unknown (anti-metabolite, inhibits dihydrofolic acid reductase)	Yes
Mycophenolate mofetil	Non-biologic	CellCept	Guanosine synthesis inhibitor	No
Sulfasalazine	Non-biologic	Azulfidine Sulfazine	Unknown	Yes
Tacrolimus (FK506)	Non-biologic	Prograf	Reduces T-cell and IL-2 activity	No
Thalidomide	Non-biologic	Thalomid	Unknown	No

Abbreviations: CD = cluster of differentiation; Fc = fragment crystallizable; FDA = U.S. Food and Drug Administration; IgM = immunoglobulin M; IL = interleukin; IVIG = intravenous immunoglobulin; JIA = juvenile idiopathic arthritis; T-cell/-lymphocyte = thymus cell/lymphocyte; TNF = tumor necrosis factor

Table 2. Conventional treatments evaluated

Generic name	Drug type	U.S. trade name(s)	FDA-approved for JIA?
Betamethasone	Intra-articular corticosteroid	Celestone	Yes
Triamcinolone acetonide	Intra-articular corticosteroid	Kenolog	Yes
Triamcinolone hexacetonide	Intra-articular corticosteroid	Aristospan	No
Celecoxib	NSAID	Celebrex	Yes
Etodolac	NSAID	Lodine	No
Ibuprofen	NSAID	Motrin, Advil	Yes
Indomethacin	NSAID	Indocin, Indocin SR	Yes
Meloxicam	NSAID	Mobic	Yes
Naproxen	NSAID	Naprosyn, Aleve	Yes
Oxaprozin	NSAID	Daypro	Yes
Tolmetin	NSAID	Tolectin	Yes

Abbreviations: FDA = U.S. Food and Drug Administration; JIA = juvenile idiopathic arthritis; NSAID = non-steroidal anti-inflammatory drug

Methods

Topic Refinement

The topic for this report was nominated in a public process. With input from a group of key informants, the topic was refined to assure its relevance to stakeholders, after which the proposed analytic framework and key questions were posted to a public website for comment. The EPC subsequently revised the analytic framework and key questions based on the comments received.

Search Strategy

We conducted a comprehensive search of the scientific literature to identify randomized controlled trials (RCTs), nonrandomized comparative studies, case series, and case reports relevant to the key questions. Searches of electronic databases used the National Library of Medicine's Medical Subject Headings (MeSH) keyword nomenclature developed for MEDLINE®. For Key Questions 1-4 we combined search terms for JIA with terms for the interventions of interest; for Key Question 5 we supplemented this basic search with general terms for clinical outcomes measures and specific terms for the measures of interest. Detailed search strategies are provided in Appendix A. We also reviewed selected gray literature identified by the SRC, the reference lists of relevant review articles, and abstracts presented at relevant meetings (the 2008 and 2009 meetings of the American College of Rheumatology and the 2008 and 2009 meetings of the Pediatric Academic Societies) for all key questions.

To identify literature describing the comparative benefits and harms of DMARDs (Key Questions 1-4) and the accuracy of clinical outcome measures (Key Question 5) we searched:

- MEDLINE® via PubMed (1996 to January 31, 2010);
- The Cochrane Database of Systematic Reviews;
- Gray literature identified by the SRC.

For Key Question 3, we also searched the FDA Adverse Event Reporting System from July 2008 to June 2009.

Our searches identified a total of 2072 citations. We imported all citations into an electronic database (EndNote® version X13).

Study Selection

We developed criteria for inclusion and exclusion based on the patient populations, interventions, and outcome measures specified in the key questions. The abstract screening criteria we applied are listed in Appendix B. We then applied a second, more stringent set of criteria for inclusion and exclusion at the full-text stage (Appendix B). In general, we included peer-reviewed, English-language reports of studies that had a sample population of individuals 18 years or younger with JIA according to the current ACR definition. For Key Questions 1, 2,

and 4, the study duration had to be at least 3 months. To be included for Key Questions 1-4, studies had to include at least one of the DMARDs included in our list. In addition, Key Questions 1 and 2 each required comparators. For Key Question 1, the comparator was conventional treatment, and for Key Question 2, the comparator was another DMARD. Case reports could be included for Key Question 3. For Key Question 5, any treatment intervention or comparator (including none) and any study duration were acceptable. We restricted Key Question 5 to studies of specified clinical outcome measures for childhood JIA.

The remainder of this section describes in greater detail the criteria we used to screen the available literature.

Population and Condition of Interest

This review focused on individuals aged 18 years or younger with:

- Juvenile idiopathic arthritis (JIA) according to the International League of Associations for Rheumatology (ILAR) criteria; or
- Juvenile rheumatoid arthritis (JRA) according to the American College of Rheumatology (ACR) definition; or
- Juvenile chronic arthritis (JCA) according to the European League Against Rheumatism (EULAR) criteria.

Any diagnostic subtype of any severity was acceptable. In many cases, insufficient information was reported to verify the diagnosis; therefore, we accepted diagnoses as reported by the study authors. We included studies with patients of mixed ages only if results were reported separately for the relevant subgroups.

Interventions and Comparators of Interest

For Key Questions 1, 2, and 4, we included DMARDs as listed in Table 1 as the interventions of interest. The comparator was conventional treatment (Table 2), defined as NSAIDs or intra-articular corticosteroids with or without methotrexate. Many studies evaluated DMARDs plus conventional treatment versus conventional treatment alone. We considered methotrexate to be a component of the test intervention if the comparator group did not receive methotrexate. We considered methotrexate to be a component of the comparator if individuals in both the treatment and comparison groups could receive methotrexate. Key Questions 3 and 5 did not require a comparator.

Outcomes of Interest

We considered a wide range of outcomes pertaining to the benefits and harms of DMARDs (Key Questions 1-4) and the utility of clinical outcome measures (Key Question 5). These outcomes included:

For Key Questions 1-4:

- Efficacy outcomes: Improvement in intermediate or long-term outcomes. Intermediate outcomes included laboratory measures of inflammation, active joint

count, number of joints with limited range of motion, radiographic evidence of the progression of disease, and global assessment of current status. Long-term outcomes included pain control, clinical remission, quality of life, growth, development, joint function, functional ability, and mortality.

- Adverse events: These are specific to the interventions being examined. Because of the known risks associated with DMARDs, we focused primarily on serious infections and the development of cancer when assessing adverse events. Other categories we examined included mortality, hepatitis, bone marrow suppression, nausea or vomiting, and risks to fetus or pregnant mother.

For Key Question 5:

- Outcomes of interest: Inter- and intra-rater reliability, test-retest reliability, responsiveness (standardized response mean and responsiveness index), time to administer, and construct validity.

  Instruments evaluated: Based on studies identified in our search for articles relevant to Key Questions 1-4, and in consultation with the project's technical expert panel (TEP), we selected for detailed review the instruments most commonly used in clinical trials and newer instruments of growing importance. These included: measures of disease activity (active joint count, physician global assessment of disease activity, parent/patient global assessment of well-being), a measure of functional status/disability (Childhood Health Assessment Questionnaire), measures of health-related quality of life (Child Health Questionnaire, Pediatric Quality of Life Inventory 4.0, Pediatric Quality of Life Inventory Rheumatology Module), and composite measures of disease status or response to therapy (American College of Rheumatology Response Criteria, remission, flare, minimal disease activity). We chose to focus on studies in which the instrument's psychometric characteristics were examined specifically for children with JIA. Therefore, we excluded initial psychometric evaluations of general health-related quality-of-life instruments conducted in children without JIA and studies of disease-specific instruments in which children with JIA were only a small proportion of the overall sample.

Timing

We included comparative studies that evaluated the efficacy or effectiveness of treatment if the intervention period lasted at least 3 months (Key Questions 1, 2, or 4). We included all reports of adverse events, regardless of the duration of treatment (Key Question 3). We also included all studies of clinical outcomes measures (Key Question 5), regardless of followup duration.

Setting

We did not restrict the setting of the included studies.

Types of Studies

To evaluate the efficacy or effectiveness of treatment and adverse events (Key Questions 1-4), we included prospective comparative clinical studies of any design, including randomized controlled trials (RCTs), nonrandomized controlled clinical trials, and cohort studies. To evaluate adverse events (Key Question 3), we also included case series and case reports. To evaluate clinical outcomes measures (Key Question 5), we considered prospective clinical studies and cross-sectional studies.

Analytic Framework

Figure 1 depicts the key questions within the context of the population, interventions, comparators of interest, outcomes, timing, and settings (PICOTS). In general, the figure illustrates how treatment of JIA in children with DMARDs versus conventional treatment (intra-articular corticosteroids and NSAIDs with or without methotrexate) may result in intermediate outcomes, such as changes in laboratory measures of inflammation, changes in the active joint count, or radiographic progression of disease, and/or long-term outcomes, such as clinical remission, changes in quality of life, changes in growth, and changes in development. Also, adverse events may occur at any point after the treatment is received.

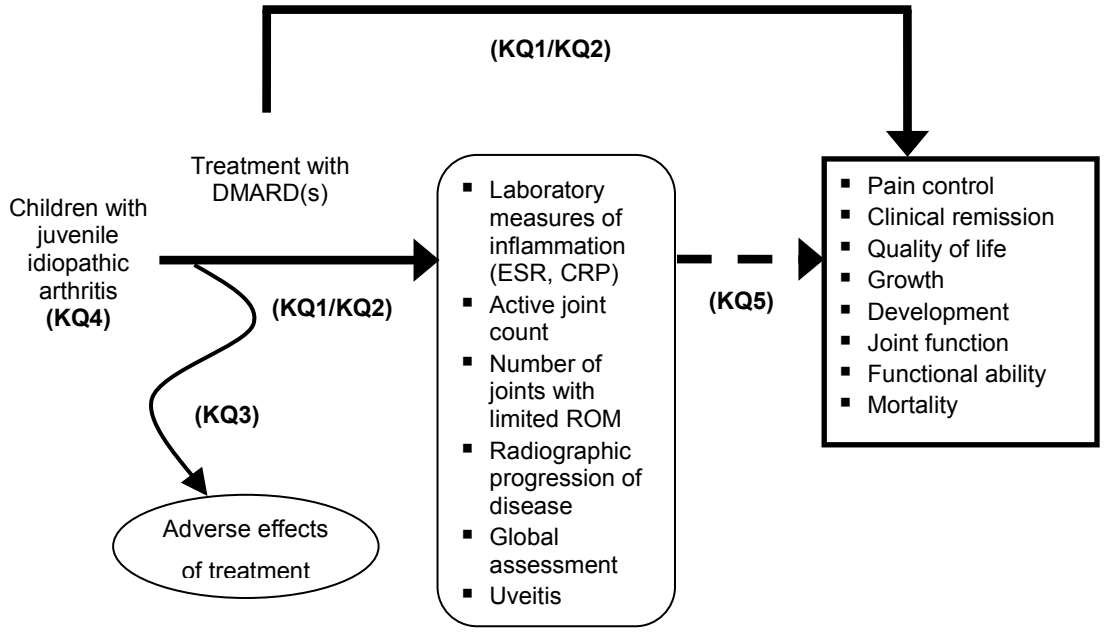

Figure 1. Analytic framework

Data Extraction

We developed separate data abstraction form/evidence table templates for abstracting data from included studies that addressed treatment effects (benefits and adverse effects) and the performance of clinical outcome instruments (Appendix C). Abstractors worked in pairs: the first abstracted the data, and the second over-read the article and the accompanying abstraction to check for accuracy and completeness. Completed evidence tables are provided in Appendix D.

For studies reporting efficacy outcomes, we extracted the following data from clinical trials and cohort studies: geographical location; study dates; funding source; interventions (including dose, duration, dose titration protocol [if any], and cointerventions [if any]); study design; population characteristics (including age, sex, race/ethnicity, type of JIA, baseline severity, and comorbidities); recruitment setting; inclusion and exclusion criteria; numbers screened, eligible, enrolled, and lost to followup; and results for each outcome.

For adverse events, we also abstracted data from case series and case reports. We developed an Excel spreadsheet to abstract the following data from both the peer-reviewed, published literature, as well as the gray literature, including published abstracts and letters to the editor: DMARD interventions, study design, total sample size, intervention sample size, gender, and the nature of the adverse event. There was wide variability across studies in how adverse events were defined, ascertained, and reported, and different terms were used to report similar events (e.g., "rash," "skin changes," "dermatitis," or "dermatologic event"). To facilitate comparisons across studies and interventions for the purpose of this report, we developed a classification system that included 29 categories (including death), plus an "other" category. Patients who experienced multiple different adverse events thus contributed data points to the respective adverse event categories. We did not abstract multiple symptoms for a given patient when these symptoms were all attributed by the authors to a given diagnosis (e.g., a patient diagnosed with pneumonia and reporting symptoms of cough, fever, chest pain, and dyspnea contributed only to the "respiratory" adverse event category). We included a given diagnosis only once (e.g., we classified "pneumonia" as a

respiratory adverse event rather than "infection"). A single investigator abstracted, categorized, and summarized the adverse events data for this report. Results are given in Appendix E.

Quality Assessment

For Key Questions 1, 2, and 4, we used predefined criteria to assess the quality of individual controlled trials and prospective cohort studies. For this purpose, we adapted criteria from AHRQ's Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews.[5]

Individual studies were graded as "good," "fair," or "poor" in quality according to the following definitions:

> A "good" study has the least bias and results are considered valid. A good study has a clear description of the population, setting, interventions, and comparison groups; uses a valid approach to allocate patients to alternative treatments; has a low dropout rate; and uses appropriate means to prevent bias, measure outcomes, and analyze and report results.
>
> A "fair" study is susceptible to some bias, but probably not sufficient to invalidate the results. The study may be missing information, making it difficult to assess limitations and potential problems. As the fair-quality category is broad, studies with this rating vary in their strengths and weaknesses. The results of some fair-quality studies are *possibly* valid, while others are *probably* valid.
>
> A "poor" rating indicates significant bias that may invalidate the results. These studies have serious errors in design, analysis, or reporting; have large amounts of missing information; or have discrepancies in reporting. The results of a poor-quality study are at least as likely to reflect flaws in the study design as to indicate true differences between the compared interventions.

If a study was rated as fair or poor, assessors were instructed to note important limitations on internal validity related to the following variables:

1) Initial assembly of comparable groups.
2) Maintenance of comparable groups (includes attrition, crossovers, adherence, and contamination).
3) Important differential loss to followup or overall high loss to followup.
4) Measurements: Equal, reliable, and valid (includes masking of outcome assessment).
5) Clear definition of interventions.
6) All important outcomes considered.
7) Analysis: Adjustment for potential confounders for cohort studies, or intention-to-treat analysis for RCTs.

Assessment of each study's quality was made by a single rater and then evaluated by a second rater. Disagreements were resolved by consensus. Final quality assessments for individual studies are included in the evidence tables (Appendix D).

Quality was not rated for the case reports and case series included for Key Question 3. No established quality measurement evaluation systems have been developed for studies evaluating the reliability and validity of clinical outcome measures (Key Question 5). We therefore adapted pertinent criteria from the QUADAS tool used to assess the quality of diagnostic tests studies.[6] We considered the selection of study participants, independent and blind comparison of the study instrument to other outcome measures, and the appropriateness of the analytic approach.

Rating the Body of Evidence

We assessed the strength of the body of evidence for each key question using the a modified version of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.[7] Unlike GRADE, the EPC GRADE method does not make specific clinical recommendations, and "insufficient" is used instead of "very low" when an estimate of effect cannot be generated.[5] In rating the strength of evidence, we considered the number of studies, the size of the studies, strength of study design, and the quality of individual studies. In addition, as part of the GRADE framework, we assessed the consistency across studies of the same design, consistency across different study designs, the magnitude of effect, and applicability. Finally, if applicable, we considered the likelihood of publication bias and (especially for observational studies) the potential influence of plausible confounders. We commented specifically when it was difficult or impossible to assess certain of these dimensions. The overall strength of a given body of evidence was rated qualitatively using the following four-level scale:

> High – High confidence that the evidence reflects the true effect. Further research is very unlikely to change our confidence in the estimate of effect.
>
> Moderate – Moderate confidence that the evidence reflects the true effect. Further research may change our confidence in the estimate of effect and may change the estimate.
>
> Low – Low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate.
>
> Insufficient – Evidence either is unavailable or does not permit estimation of effect.

Assessing Applicability

We followed the recommendations in AHRQ's *Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews*[5] by abstracting data on the population studied, the intervention and comparator, the outcomes measured and timing of assessments. We used these data to evaluate the applicability to clinical practice, paying special attention to study eligibility

criteria, symptom severity and subtypes of JIA for the included sample, DMARD dose and comparators, and clinical relevance and timing of the outcome measures. Using notations on applicability from the evidence tables along with our summary tables, we summarized issues of applicability qualitatively.

Data Synthesis

We planned to perform meta-analysis if there were sufficient studies that were conceptually homogeneous and reported the needed data to compute a summary estimate. In deciding whether to conduct meta-analyses, we considered primarily the basic study design (e.g., RCT), the intervention, and the comparator. Because of the small number of included studies and heterogeneity in comparisons, meta-analysis was conducted for only one comparison; all other literature was synthesized qualitatively. Meta-analysis was performed using Review Manager, version 5.0.24.[8] The pooled effects estimate for the binary outcome was expressed as a risk ratio (RR) with 95 percent confidence interval (CI). We tested the difference in estimates of treatment effect between the treatment and control groups using a 2-sided z test with statistical significance considered at a P value of less than 0.05. We examined heterogeneity by using the Cochran Q and the I^2 test.[9,10] We predefined heterogeneity as low, moderate, and high, with I^2 statistics greater than 25 percent, 50 percent, and 75 percent, respectively.[9] Meta-analysis with a fixed-effect model was utilized because the observed heterogeneity was low.[10] For Key Question 3, we used results from clinical trials and cohort studies to describe rates of adverse effects. We used case reports and case series to describe potential adverse events that have not been reported in clinical trials.

Peer Review Process

Peer review was conducted to provide independent evaluation of the systematic review methods and content. External stakeholders nominated to review this report included clinicians and representatives of professional societies, as well as members of the TEP. AHRQ concurred with these nominees to conduct peer review based on an assessment of their independence and expertise. The review is also available for public comment by other stakeholders and experts.

Results

Literature Search and Screening

Searches of all sources identified a total of 2072 potentially relevant citations. Table 3 details the number of citations identified from each source.

Table 3. Sources of citations

Source	Number of citations
MEDLINE®	1732
Gray literature identified by the SRC	314
References of review articles and primary studies	11
Other (recommendations from staff at AHRQ or TEP or from project investigators)	13
Total:	**2072**

Figure 2 describes the flow of literature through the screening process. Of the 2072 citations identified by our searches, 1348 were excluded at the abstract screening stage. Of the 724 articles that passed the initial abstract screening, 303 were gray literature articles that were reviewed separately and excluded from further review. The remaining 421 articles went on to full-text screening. Of these, 255 were excluded, leaving a total of 166 included articles. Appendix F provides a complete list of articles excluded at the full-text screening stage, with reasons for exclusion.

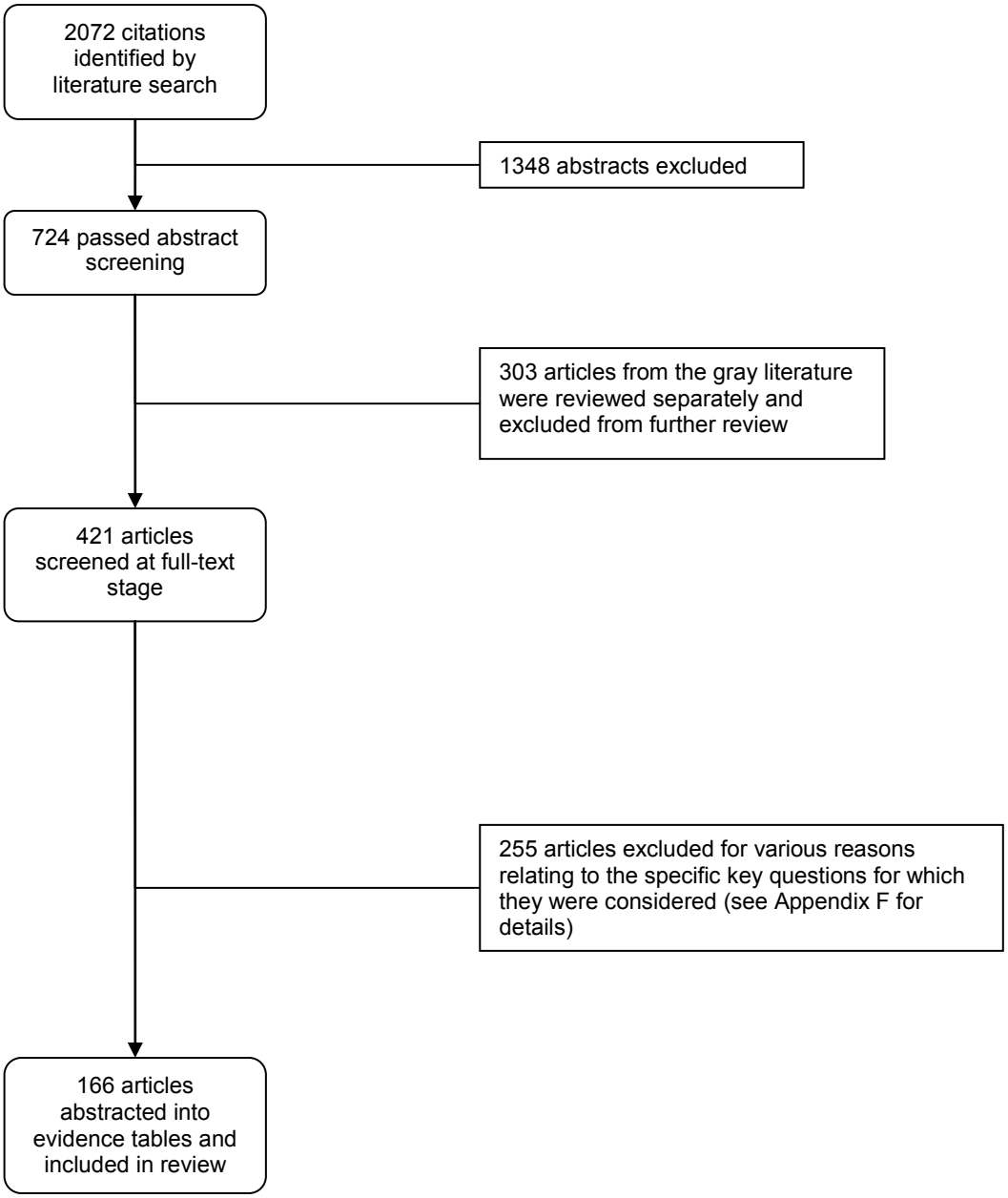

Figure 2. Literature flow diagram

Figure 3 summarizes the treatment comparisons evaluated in the included efficacy studies (Key Questions 1, 2, and 4). Six non-biologic DMARDs and seven biologic DMARDs have been compared to conventional treatment with or without methotrexate. Two different sets of non-biologic DMARDs have been directly compared (leflunomide vs. methotrexate and hydroxychloroquine vs. penicillamine), and two biologic DMARDs have been directly compared (etanercept vs. infliximab). Three of the biologic DMARDs that have been compared to conventional treatment were in the same class (TNF inhibitors: adalimumab, etanercept, and

infliximab). However, study heterogeneity precluded meta-analysis of this combined class versus conventional treatment. Details on the number of studies describing each treatment comparison are provided under the relevant Key Question, below.

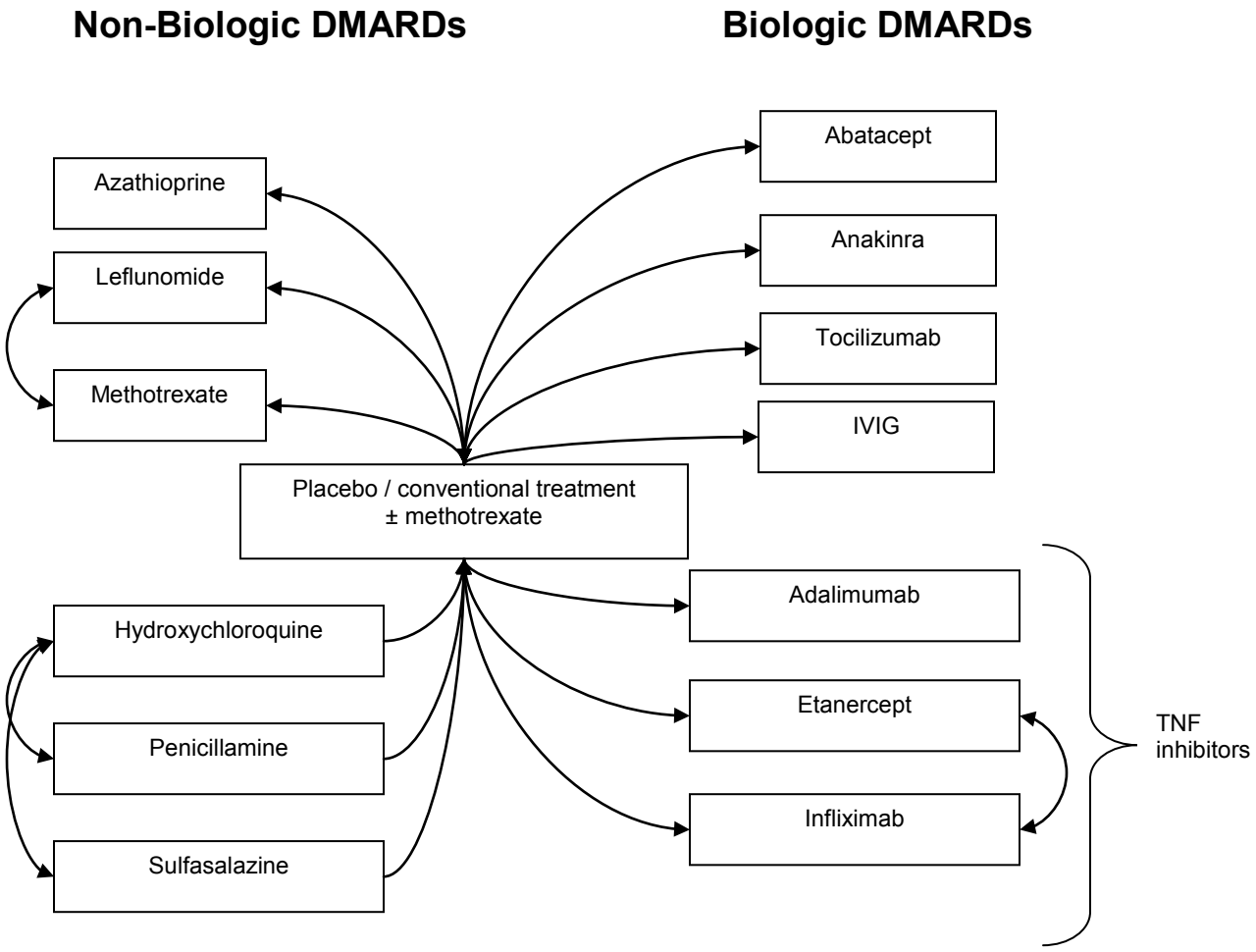

Figure 3. Treatment comparisons evaluated in efficacy studies

Key Question 1. In children with JIA, does treatment with DMARDs, compared to conventional treatment, improve laboratory measures of inflammation or radiological progression, symptoms (e.g., pain, symptom scores), or health status (e.g., functional ability, mortality)?

Key Points

- Among the non-biologic DMARDs, there is some evidence that methotrexate is superior to conventional therapy and oral corticosteroids.
- Among children who have responded to a biologic DMARD, randomized discontinuation trials suggest that continued treatment for 4 months to 2 years decreases the risk of having a flare. Although these studies evaluated DMARDs with different mechanisms of action (abatacept, adalimumab, anakinra, etanercept, intravenous immunoglobulin [IVIG], tocilizumab) and used varying comparators, followup periods, and descriptions of flare, the finding of a reduced risk of flare was precise and consistent.
- Conventional treatment has changed over time (e.g., use of oral corticosteroids in older studies of non-biologic DMARDs versus more frequent use of methotrexate in more recent studies of biologic DMARDs). Comparing the effectiveness of biologic and non-biologic DMARDs is challenging because of variations in comparators and how these comparators are described.
- There is significant variation in outcome measures and how these outcome measures are reported.

Detailed Analysis

Literature Identified

We identified of 20 publications describing 18 unique studies and involving 1532 patients that compared DMARDs to conventional treatments with or without methotrexate. Among these were 10 studies that evaluated seven biologic DMARDs (abatacept, adalimumab, anakinra, etanercept, infliximab, IVIG, and tocilizumab; see Table 4) and eight studies that evaluated five non-biologic DMARDs (azathioprine, penicillamine, hydroxychloroquine, methotrexate, and sulfasalazine; see Table 5).

There were 10 RCTs, of which four (described in five papers) were of good quality,[11-15] four were of fair quality,[16-19] and two were of poor quality.[20,21] Key problems in the fair- and poor-quality studies included unclear methods of allocating to therapy, questionable blinding, and incomplete followup. There were two open-label comparison studies of poor quality.[22,23] Six studies were randomized discontinuation studies, of which three (described in four papers) were of good quality,[24-27] two were of fair quality,[28,29] and one was of poor quality.[30]

A detailed summary of these studies, by DMARD evaluated, is provided below.

There were no good-quality RCTs comparing biologic DMARDs to conventional therapy. There were two good-quality RCTs comparing methotrexate, a non-biologic DMARD, to conventional therapy.[13,14] However, in both studies, each group could also receive oral corticosteroids, which are not currently considered conventional therapy. A single good-quality trial of sulfasalazine showed better short-term (24-week) outcomes than treatment with NSAIDs.[15]

Table 4. Studies comparing biologic DMARDs versus conventional treatments with or without methotrexate

Study	DMARD(s)	Comparator(s)	Other arthritis drugs	Study design	Study quality	Study population (n)	Followup duration	Key questions addressed
Ruperto et al., 2008[24]	Abatacept	Placebo	Methotrexate, oral corticosteroids, NSAIDs, analgesics	Randomized discontinuation trial with open label followup	Good	JIA: - Persistent oligoarthritis (5) - Extended oligoarthritis (43) - Polyarthritis (205) - Systemic (60)	6 months (RCT) with 5-yr followup	1, 3
Lovell et al., 2008[25]	Adalimumab	Placebo	Methotrexate, NSAIDs, oral corticosteroids	Randomized discontinuation trial with open label followup	Good	JRA - Polyarticular (171)	32 weeks (RCT) up to 56 week followup	1, 3
Ilowite et al., 2009[30]	Anakinra	Placebo	Methotrexate, NSAIDs, oral corticosteroids	Randomized discontinuation trial with open label followup	Poor	JIA: - Polyarticular (33) - Pauciarticular (6) - Systemic (11)	16 week (RCT) and 12 month followup	1, 3, 4
Lovell et al., 2000[26] and Lovell et al., 2003[27]	Etanercept	Placebo	NSAIDs, oral corticosteroids, pain medication except for 12 hours before joint assessment	Randomized discontinuation trial with open label followup	Good	JRA: - Polyarticular (62) - Pauciarticular (6) - Systemic (34)	4 month (RCT) and 2-year followup	1, 3, 4
Smith et al., 2005[16]	Etanercept	Placebo	Methotrexate, prednisone	RCT	Fair	JRA with uveitis (12)	12 months	1, 3
Ruperto et al., 2007[17]	Infliximab or infliximab with methotrexate	Placebo and methotrexate	NSAIDs, opioids, oral corticosteroids	RCT with active treatment extension	Fair	JRA - Polyarticular onset (74) - Pauciarticular onset (28) - Systemic onset (19)	52 weeks	1, 3
Giannini et al., 1996[28]	IVIG	Placebo	NSAIDs, methotrexate, sulfasalazine, hydroxy-	Randomized discontinuation trial with open label followup	Fair	JRA - Polyarticular (19)	4 months (RCT)	1, 3

Study	DMARD(s)	Comparator(s)	Other arthritis drugs	Study design	Study quality	Study population (n)	Followup duration	Key questions addressed
			chloroquine					
Oppermann et al., 1994[22]	IVIG	Methyl-prednisolone	NSAIDs, methotrexate, oral corticosteroids	Open-label comparison	Poor	JCA (20)	Unclear; 6-8 months?	1
Silverman et al., 1994[20]	IVIG	Placebo	NSAIDs, up to 2 SAARDs (not listed)	RCT	Poor	JRA - Systemic (31)	6 months	1, 3
Yokota et al., 2008[29]	Tocilizumab	Placebo	Oral corticosteroids	Randomized discontinuation trial with open label followup	Fair	JIA (43)	12 week RCT, 48 week followup	1, 3

Abbreviations: DMARD(s) = disease-modifying anti-rheumatic drug(s); JCA = juvenile chronic arthritis; JIA = juvenile idiopathic arthritis; JRA = juvenile rheumatoid arthritis; NSAID(s) = non-steroidal anti-inflammatory drug(s); RCT = randomized controlled trial; SAARD(s) = slow-acting anti-rheumatic drug(s)

Table 5. Studies comparing non-biologic DMARDs versus conventional treatments with or without methotrexate

Study	DMARD(s)	Comparator(s)	Other arthritis drugs	Study design	Study quality	Study population (n)	Followup duration	Key questions addressed
Kvien et al., 1986[18]	Azathioprine	Placebo	NSAIDs, prednisolone	RCT	Poor	JRA: - Polyarticular-onset (16) - Pauciarticular onset (9) - Systemic onset (7)	16 weeks	1, 3
Prieur et al., 1985[19]	Penicillamine	Placebo	Pyridoxine hydrochloride	RCT	Fair	JCA: - Polyarticular onset (35) - Pauciarticular onset (14) - Systemic onset (25)	6 months	1, 3
Brewer et al., 1986[11] and Van Kerckhove et al., 1988[12]	Penicillamine or hydroxy-chloroquine	Placebo	NSAIDs, acetaminophen, codeine	RCT	Good	JRA: - Polyarticular (142) - Pauciarticular (11) - Systemic (9)	12 months	1, 2, 3
Kvien et al., 1985[21]	Penicillamine or hydroxy-chloroquine	Gold	Acetaminophen, NSAIDs	Open-label RCT	Poor	JRA: -Polyarticular (49) -Pauciarticular (23)	50 weeks	
Riddle et al., 2006[23]	Methotrexate	NSAIDs, methylpredni-solone	Not reported	Open-label comparison	Poor	JIA (63)	4 months	1, 3
Giannini et al., 1992[13]	Methotrexate	Placebo	NSAIDs, prednisolone	RCT	Good	JIA (127)	6 months	1, 3, 4
Woo et al., 2000[14]	Methotrexate	Placebo	Prednisolone, intra-articular corticosteroids, NSAIDs	RCT with crossover	Good	JIA - Extended oligoarticular (43) - Systemic (45)	12 months	1, 3, 4
van Rossum et al., 1998[15]	Sulfasalazine	Placebo	NSAIDs	RCT	Good	JCA: - Polyarticular (32) - Oligoarticular (37)	24 weeks	1, 3

Abbreviations: DMARD(s) = disease-modifying anti-rheumatic drug(s); JCA = juvenile chronic arthritis; JIA = juvenile idiopathic arthritis; JRA = juvenile rheumatoid arthritis; NSAID(s) = non-steroidal anti-inflammatory drug(s); RCT = randomized controlled trial

Biologic DMARDs Versus Conventional Treatment With or Without Methotrexate

Abatacept

One good-quality randomized discontinuation study evaluated abatacept.[24] During the 6-month double-blind period of this study, there was statistically significant improvement compared to placebo in the active joint count (4.4 vs. 6; $p = 0.02$), CHAQ score (0.8 vs. 0.7; $p = 0.04$), physician global assessment (14.7 vs. 12.5; $p < 0.01$), and ACR Pediatric 90 (40 percent vs. 16 percent; $p < 0.01$). There was no statistically significant improvement in parent/patient global assessment (17.9 vs. 23.9; $p = 0.70$) or erythrocyte sedimentation rate (ESR; 25.1 vs. 30.7; $p = 0.96$).

Adalimumab

We found one good-quality randomized discontinuation trial that compared adalimumab to conventional therapy.[25] The results were stratified by use of methotrexate. At the end of the 48-week double-blind phase, the proportion of patients who had a flare of disease in the adalimumab without methotrexate group was lower than in the conventional treatment group without methotrexate (43 percent vs. 71 percent; $p = 0.03$), and lower than in those groups that did receive methotrexate (37 percent vs. 65 percent; $p = 0.02$). The proportion who achieved ACR Pediatric 50 score in the adalimumab without methotrexate group was higher than in the conventional treatment without methotrexate group (53 percent vs. 32 percent; $p = 0.01$), and higher than in those groups that received methotrexate (63 percent vs. 38 percent; $p = 0.03$). Although the proportion who achieved ACR Pediatric 90 score was higher in the adalimumab without methotrexate group than in the conventional treatment without methotrexate group (30 percent vs. 18 percent), the difference was not statistically significant ($p = 0.28$). Similarly, the difference in the proportion who achieved the ACR Pediatric 90 among those who also received methotrexate was higher in the adalimumab group than in the conventional treatment group, but did not achieve statistical significance (42 percent vs. 27 percent; $p = 0.17$).

Anakinra

One randomized discontinuation trial compared anakinra to conventional therapy.[30] This study was rated as poor in quality because it did not have sufficient statistical power to evaluate efficacy, there was insufficient reporting of randomization and concealment, and there was a potential conflict of interest because of the source of funding. The main goal of the study was to evaluate safety. By week 28 of blinded treatment, 16 percent who received anakinra and 40 percent who received placebo had had a flare ($p = 0.11$). There was improvement in the CHAQ score in the anakinra group compared to placebo (-0.25 vs. 0.13; no p-value reported). Similarly, there was improvement in the ESR among those who were treated with anakinra (-2.21 vs. 13.73; no p-value reported).

Etanercept

Two studies evaluated etanercept versus placebo. One good-quality randomized discontinuation trial evaluated only children with polyarticular JRA.[26] In the double-blind component, fewer

patients who received etanercept had a flare (28 percent vs. 81 percent; p = 0.003). There was also an improvement in the CHAQ score (-0.8 vs. -0.1). Overall, there was a 54 percent median improvement among those who received etanercept compared to no median change in the placebo group. There was an overall improvement in the number of active joints (7 vs. 13; no p-value reported); physician global assessment (2 vs. 5; no p-value reported); parent global assessment (3 vs. 5; no p-value reported); ESR (18 vs. 30; no p-value reported); and the proportion who achieved ACR Pediatric 50 (72 percent vs. 23 percent; no p-value reported).

The other study of etanercept was a fair-quality RCT that evaluated effectiveness for the treatment of uveitis.[16] This study had a small sample size, and there was potential for a conflict of interest with the study sponsor. During the study, 6 of 12 in the test treatment arm and 2 of 5 in the conventional treatment arm improved. This was described by study investigators as no apparent difference.

Infliximab

One fair-quality RCT compared infliximab to conventional treatment.[17] This study inconsistently and incompletely reported outcomes. The study did not find statistically significant differences between infliximab and conventional treatment in the ACR Pediatric 50 at 14 weeks (50 percent vs. 33.9 percent, respectively; p = 0.13) or the rate of clinical remission at 52 weeks (44.1 percent vs. 43.1 percent, respectively).

IVIG

Three studies compared IVIG to conventional treatment. One small (19 total in the double-blind phase), fair-quality, randomized discontinuation trial[28] found a 3 percent decrease in the active joint count among those who were treated compared to a 30 percent increase in the placebo group. Physician global assessment improved for 3 percent of patients in the treatment group and worsened for 91 percent in the placebo group. This study used a main outcome measure that has not been validated and provided no statistical significance testing; there was also a potential conflict of interest with the study sponsor.

Another study[22] compared IVIG to methylprednisolone. This study was considered to be of poor quality because it was open-label and non-randomized, analyses were not adjusted for baseline differences, and the sample was not adequately described. Investigators found no statistically significant difference between the IVIG and methylprednisolone groups for ESR (59 at baseline and 21 at 6 months vs. 61 at baseline and 24 at 6 months, respectively).

A small RCT[20] found that IVIG compared to conventional therapy was associated with a non-statistically significant improvement in the median change in active joint count (-2 vs. -1) and in physician global assessment of improvement (50 percent improvement vs. 27 percent improvement; p > 0.3). This study was considered to be of poor quality because of the small sample size and high dropout rate.

Tocilizumab

One fair-quality randomized discontinuation trial evaluated tocilizumab.[29] The screening and randomization procedures were not described. There was potential for significant conflict of interest because the data were analyzed by the study sponsor, which had a financial interest in tocilizumab. No p-values were reported for the outcomes of interest in this review. From the RCT component, the active joint count in the tocilizumab group decreased from 3.5 to 0. Similarly, in the conventional treatment group it decreased from 4 to 0. There was improvement in the CHAQ score for each group (-0.5 vs. -0.25). Both physician global assessment (51.0 to 5.5 vs. 51 to 14) and parent global assessment (51.0 to 4.5 vs. 55 to 39) improved. The ESR decreased for both the tocilizumab and conventional treatment group (35 to 0.1 vs. 38 to 15). The ACR Pediatric scores were reported graphically. The ACR Pediatric 70 increased in the tocilizumab group from approximately 70 percent to approximately 80 percent, but decreased in the conventional treatment group from approximately 80 percent to approximately 30 percent.

Meta-analysis of randomized discontinuation trials

Randomized discontinuation trials include only patients who initially responded to a treatment and primarily assess the risk of worsening when treatment is withdrawn. These studies evaluate sustainability of treatment effects and not the potential treatment effect among those who have not yet begun treatment. The randomized discontinuation trials identified by our search evaluated only biologic DMARDs (abatacept, adalimumab, anakinra, etanercept, IVIG, tocilizumab).

Four of the trials reported flare of arthritis,[24-26,30] allowing us to calculate a summary measure of the risk of flare over the 4-month to 2-year durations of the studies. Other outcomes were too heterogeneous or were reported too incompletely to calculate a summary estimate. Although there were differences in the interventions, comparators, and duration of followup among the four studies, we found very little statistical heterogeneity. Figure 4 summarizes the risk ratio (RR) for flare (with 95 percent confidence interval [CI]). Overall, the RR for having a flare among those who continued compared to those who discontinued was 0.46 (95 percent CI 0.36 to 0.60) over 4 months to 2 years. Although there is heterogeneity in study design, the RR for having a flare was similar across all studies ($x^2 = 3.18$, df = 3, p = 0.36; $I^2 = 6$ percent). This suggests that among those who respond to a biologic DMARD, there is a significant risk of flare after discontinuation. There was insufficient evidence regarding the efficacy of the biologic DMARDs from the other studies that compared these treatments to conventional therapy with or without methotrexate.

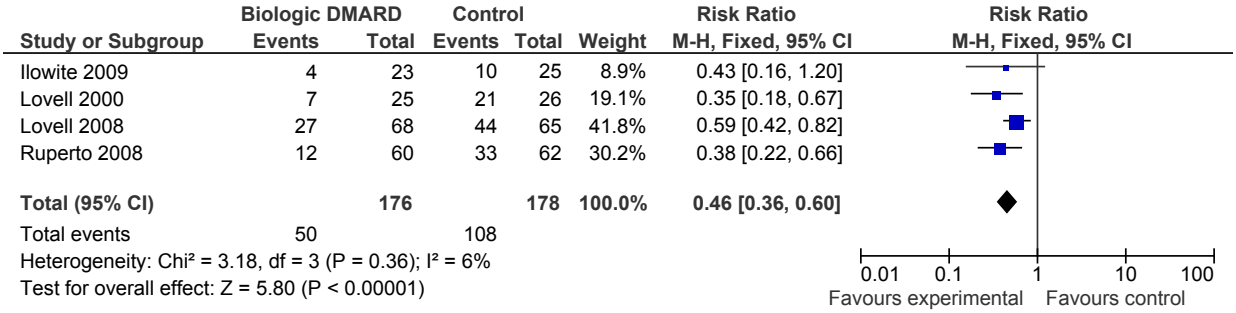

Figure 4. Comparison of symptomatic flare in children with JIA randomized to continuing a biologic DMARD versus placebo

Non-Biologic DMARDs Versus Conventional Treatment With or Without Methotrexate

Azathioprine

One poor-quality RCT evaluated azathioprine.[18] Allocation was not specified; there were baseline differences between those who received and did not receive azathioprine; it was unclear if outcomes were assessed blinded to the intervention status of subjects; and the outcomes were not well described. At 16 weeks of treatment, this study found non-statistically significant improvements with azathioprine in the number of active joints (-7 vs. -1; p = 0.45), physician global assessment (-5 vs. -2; p = 0.12), and the proportion with 50 percent improvement in ESR (4/13 subjects vs. 2/11 subjects; p = 0.36).

Hydroxychloroquine

Two RCTs evaluated hydroxychloroquine. One (described in two publications[11,12]) found no significant difference in the change in mean active joint count compared to placebo after 12 months (6.7 [95 percent CI -9.4 to -4] vs. -5.4 [-8 to -2.8]). The physician global assessment appeared slightly better for hydroxychloroquine than for placebo (70 percent better, 26 percent same, 2 percent worse compared to 53 percent better, 41 percent same, 6 percent worse; no p-value reported). There was no difference in the mean ESR decrease at 12 months (10 each).

The other study was an open-label RCT that compared hydroxychloroquine to gold.[21] This study was considered to be of poor quality because allocation concealment was not specified, there were important baseline differences between the treatment groups, it was unclear if outcomes were assessed blinded to the intervention, and the outcomes were not well described. At 50 weeks, there were no statistically significant differences in the active joint count (-4 vs. -5), median change in the physician global assessment (-8 vs. -9), or change in the ESR (-12 vs. -11). Similarly, the physician overall assessment of at least 50 percent improvement was not statistically significantly different between the hydroxychloroquine group and the gold group (12 of 17 improved vs. 10 of 15 improved, respectively).

Methotrexate

Three studies compared methotrexate to conventional treatment without methotrexate. One good-quality RCT compared low-dose methotrexate, very low-dose methotrexate, and placebo in a 6-month trial.[13] The mean active joint count decreased with low-dose methotrexate (-7.5), very low-dose methotrexate (-5.2), and placebo (-5.2; $p > 0.3$ overall). Physician global assessment improved with low-dose methotrexate compared to placebo ($p = 0.02$), but there was no statistically significant difference between the low-dose and very low-dose methotrexate groups for this outcome ($p = 0.06$). Based on a composite index with at least 25 percent improvement in articular score and improvement according to physicians and parents, 63 percent of those in the low-dose methotrexate group improved, compare to 32 percent in the very low-dose methotrexate group, and 36 percent in the placebo group ($p = 0.013$).

Another good-quality study[14] compared methotrexate to placebo among children with extended oligoarticular JIA or systemic JIA in a double-blind RCT with crossover. Among those with oligoarticular JIA, there was statistically significant improvement in physician global assessment ($p < 0.001$) and ESR ($p < 0.001$) with methotrexate. The change in the number of joints with synovitis (-3) did not achieve statistical significance ($p < 0.1$). Similarly, among those with systemic JIA, there was improvement in physician global assessment ($p < 0.001$), but not in ESR ($p = 0.06$) or in the number of joints with synovitis ($p = 0.06$) in patients taking methotrexate.

A poor-quality, non-randomized study compared methotrexate to NSAIDs and to methylprednisolone.[23] In this study, the active joint count improved more in the methylprednisolone group than in either the methotrexate or NSAID groups (-7.1 vs. -4 vs. -0.8, respectively; $p = 0.008$). This study, however, had confounding by indication; the analysis did not adjust for potential confounders; outcomes were not assessed blinded to the treatment condition; and patients were not blinded to their treatment assignments.

Penicillamine

Four publications describing three distinct studies evaluated penicillamine. One good-quality RCT[11,12]) found no statistically significant effect on the mean active joint count with penicillamine compared to placebo after 12 months (-3 [95 percent CI -4.8 to -1.1] vs. -5.4 [-8 to -2.8]); results were similar for physician global assessment (56 percent better, 28 percent same, 16 percent worse vs. 53 percent better, 41 percent same, 6 percent worse) and mean decrease in ESR (9.4 vs. 10).

A fair-quality RCT[19] found no statistically significant effect on ESR in a 6-month study in patients treated with penicillamine compared to conventional treatment (-18 vs. -8). However, this study did find a statistically significant decrease in the number of painful joints in patients taking penicillamine (-3 vs. -1.6; $p < 0.04$). This study was of fair quality because the patients in the placebo group may have had worse disease.

A poor-quality, open-label RCT[21] found no statistically significant effect for penicillamine compared to gold at 50 weeks in the active joint count (-2.5 vs. -5), median change in the physician global assessment (-7.5 vs. -9), change in ESR (-8 vs. -11), or the proportion of

patients who had at least a 50 percent improvement based on physician assessment (8/12 vs. 10/15).

Sulfasalazine

One good RCT evaluated sulfasalazine versus placebo.[15] In this study, it was unclear which time points were compared. However, there was statistically significant improvement with sulfasalazine in active joint count (-5.54 vs. -0.78; $p = 0.005$), physician global assessment (-1.95 vs. -0.99; $p = 0.0002$), patient/parent global assessment (-0.98 vs. -0.44; $p = 0.01$), and decrease in ESR (-0.74 vs. -0.04; $p < 0.001$). The number of improved joints by x-ray findings was not statistically significantly different (0.71 vs. 0.53).

Key Question 2. In children with JIA, what are the comparative effects of DMARDs on laboratory markers of inflammation or radiological progression, symptoms (e.g., pain, symptom scores), or health status (e.g., functional ability, mortality)?

Key Points

- There are few direct comparisons of DMARDs in children with JIA, and insufficient evidence to determine if any specific drug or drug class has greater beneficial effects.

Detailed Analysis

Literature Identified

We identified six reports describing five unique studies and involving 520 patients that directly compared various DMARDs with one another (Table 6). Among these studies were one that compared two biologic DMARDs (etanercept and infliximab) and four that compared various non-biologic DMARDs (penicillamine, hydroxychloroquine, leflunomide, methotrexate, and sulfasalazine). A detailed summary of these studies, by treatment comparison, is provided below. Of the five studies, one was an open-label, non-randomized comparison, and the rest were RCTs. However, only two of the studies were considered to be of good quality (one comparing penicillamine to hydroxychloroquine and another comparing leflunomide to methotrexate); the rest were poor in quality.

Table 6. Studies comparing various DMARDs with one another

Study	DMARD(s)	Other arthritis drugs	Study design	Study quality	Study population (n)	Followup duration	Key questions addressed
Lahdenne et al., 2003[31]	Etanercept vs. infliximab (biologics)	Methotrexate, prednisolone, cyclosporine A, sulfasalazine, intra-articular corticosteroids, NSAIDs	Open-label comparison	Poor	JIA - Polyarticular (24)	12 months	2, 3
Kvien et al., 1985[21]	Penicillamine vs. hydroxy-chloroquine	NSAIDs, prednisone	Open-label RCT	Poor	JRA: - Pauciarticular onset (41) - Polyarticular onset (31)	50 weeks	2, 3
Brewer et al., 1986[11] and Van Kerckhove et al., 1988[12]	Penicillamine vs. hydroxy-chloroquine	NSAIDs, acetaminophen, codeine, antibiotics	RCT	Good	JRA: - Polyarticular (142) - Pauciarticular (11) - Systemic (9)	12 months	1, 2, 3
Hoza et al., 1991[32]	Hydroxy-chloroquine vs. sulfasalazine	NSAIDs, prednisone	RCT	Poor	JCA: Oligoartuclar onset (13) - Polyarticular onset (23) - Systemic onset (3)	6 months	2, 3
Silverman et al., 2005[33]	Leflunomide vs. methotrexate	NSAIDs, prednisone, intra-articular corticosteroids	RCT with optional extension	Good	JRA - Polyarticular (94)	16 weeks (RCT) then 32 weeks	2, 3

Abbreviations: DMARD(s) = disease-modifying anti-rheumatic drug(s); JCA = juvenile chronic arthritis; JIA = juvenile idiopathic arthritis; JRA = juvenile rheumatoid arthritis; NSAIDs = non-steroidal anti-inflammatory drugs; RCT = randomized controlled tria

Comparisons of Biologic DMARDs

Etanercept vs. Infliximab

One poor-quality, non-randomized, open-label study compared etanercept to infliximab.[31] This study was considered to be of poor quality because drug switching made it hard to interpret findings, few data were provided about the subjects, assessment was not blinded to therapy, and no statement was made regarding the funding source. After 12 months of treatment, the change in active joint count was similar between etanercept (-9.5 [95 percent CI -19 to -3]) and infliximab (-11.5 [95 percent CI -17 to -7.5]). Results were also similar in the two treatment groups for changes in the CHAQ score (-0.81 vs. -0.31; p = 0.12), physician global assessment (-29 vs. -35; p = 0.65), patient/parent global assessment (-24.5 vs. -27.5; p = 0.81), ACR Pediatric 75 (67 percent each), and ESR (28.5 vs. -25; p = 0.37).

Comparisons of Non-Biologic DMARDs

Penicillamine vs. Hydroxychloroquine

Two publications[11,12] described a good-quality RCT that compared penicillamine and hydroxychloroquine to placebo (results described above, under Key Question 1) and to one another. At 12 months, neither active drug was superior to the other based on active joint count, ESR, or physician global assessment.

One poor-quality, open-label RCT[21] compared hydroxychloroquine and penicillamine to gold (results described above, under Key Question 1) and to one another. At 50 weeks, there were no significant differences between the two DMARDs in active joint count, physician global assessment, or ESR.

Sulfasalazine vs. Hydroxychloroquine

One poor-quality RCT compared sulfasalazine to hydroxychloroquine.[32] This study was considered to be of poor quality because there was an inadequate description of the subjects, it was unclear if the study was blinded, and many of the outcomes were not validated. After 6 months, the average number of affected joints decreased by 1.5 in the sulfasalazine group and by 0.6 in the hydroxychloroquine group (no p-value reported). During this time, the ESR decreased in both the sulfasalazine group (52.7 to 36.3; no p-value reported) and hydroxychloroquine group (41.2 to 28.9; no p-value reported). Physician global assessment (9 better, 9 worse, 3 no effect for sulfasalazine vs. 8 better, 3 worse, 7 no effect for hydroxychloroquine; no p-value reported) and patient global assessment (10 better, 7 worse, 3 no effect for sulfasalazine vs. 7 better 5 worse 3 no effect for hydroxychloroquine; no p-value reported) were similar in the two groups.

Leflumide vs. Methotrexate

One good-quality RCT compared leflumide to conventional treatment with methotrexate.[33] This 16-week study with a 32-week blinded extension found improvements in both groups. The active joint count decreased for the leflumide and conventional treatment groups (-8.1 vs. -8.9; p = not significant). Similarly, in both groups there were improvements in the CHAQ score

(-0.44 vs. -0.39; p = not significant), physician global assessment (-31.5 vs. -32.1; p = not significant), parent global assessment (-15.9 vs. -22; p = not significant), and ESR (-6.5 vs. 7.2; p = not significant). As the trial proceeded, the methotrexate group appeared to have a greater improvement in the proportion of patients who had an ACR Pediatric 30, Pediatric 50, or Pediatric 70 response. For example, 70 percent of the leflunomide group and 83 percent of the methotrexate group achieved an ACR Pediatric 70 response at 48 vs. 16 weeks. The improvement was not statistically significant for either the leflunomide (p = 0.01) or methotrexate (p = 0.06) groups. No statistical comparison was made between the two groups.

Key Question 3. In children with JIA, does the rate and type of adverse events differ between the various DMARDs or between DMARDs and conventional treatment with or without methotrexate?

Key Points

- There are few direct comparisons of DMARDs with one another in children with JIA, and insufficient evidence to determine if there are differential rates of adverse events between specific drugs or drug classes.
- Reported rates of adverse events are similar between DMARDs and placebo in nearly all published RCTs.
- Our horizon scan identified 8 incident cases of cancer among more than 3000 children treated with one or more DMARDs.
- The available data on harm must be interpreted with caution because data on adverse events have not been systematically collected or reported across studies.

Detailed Analysis

Literature Identified

Of the 15 eligible RCTs identified by our search strategy, 13 included a placebo comparison and reported adverse events. Eight of these were traditional RCTs and five were randomized discontinuation trials. Because one of these studies included three study arms, a total of 14 DMARDs or DMARD combinations were directly compared to placebo. Anakinra, abatacept, etanercept, infliximab, tocilizumab, azathioprine, hydroxychloroquine, and sulfasalazine were each represented by a single study; etanercept, IVIG, and penicillamine were each represented by two studies; and methotrexate was compared to placebo in one study and was used in combination with infliximab in another study. A total of 914 unique patients were represented in the 13 placebo-controlled trials.

Our horizon scan identified 129 publications that reported adverse events possibly associated with a DMARD among patients with JIA (Appendix E). Of these 129 publications, 19 (15 percent) were RCTs; the remainder were prospective or retrospective series or case reports. Three thousand and one hundred and thirty-three (3133) patients were represented in these

reports, with 2286 patients (73 percent) having participated in an RCT. There was insufficient information in these publications to determine if data from some patients were included in more than one published report. Furthermore, some series included some patients who were either adults or who did not have JIA.

Reporting standards for adverse events varied greatly across studies. For the purpose of this report, we consolidated the many different descriptions of reported adverse events into 24 broad categories, which we in turn categorized as involving a primary organ system, being an isolated symptom, or as "other." We did not include minor or transient events (e.g., rash) that were identified by the authors of the published reports as possibly associated with infusion of the drug.

Safety data from the 13 placebo-controlled trials are summarized in Table 7 (Parts 1-3) and described in greater detail for the specific DMARDs evaluated in the sections that follow.

Table 7 – Part 1. Adverse events reported in RCTs – dropouts and adverse events related to organ systems

DMARD	Study	Intervention	Sample size	Dropouts due to AEs	Gastrointestinal	Dermatologic	Renal/Urologic	Respiratory	Neurologic	Ophthalmologic
Biologic agents										
Abatacept	Ruperto et al., 2008[24]	Drug	62	-	-	-	-	-	-	-
		Placebo	60	-	-	-	-	-	-	-
Anakinra	Ilowite et al., 2009[30]	Drug	25	-	6	2	-	8	6	-
		Placebo	25	-	1	10	-	7	4	-
Etanercept	Lovell et al., 2000[26]	Drug	25	0	1	1				
		Placebo	26	0	-	-	-	-	-	-
	Smith et al., 2005[16]	Drug	7	-	-	-	-	-	-	-
		Placebo	5	-	-	-	-	-	-	-
Infliximab + MTX	Ruperto et al., 2007[17]	Drug + MTX	60	2	-	-	-	-	-	-
		Placebo + MTX	62	1	-	-	-	-	-	-
IVIG	Giannini et al., 1996[28]	Drug	10	0	-	-	-	-	-	-
		Placebo	9	0	-	-	-	-	-	-
	Silverman et al., 1994[20]	Drug	14	-	-	-	-	-	-	0
		Placebo	17	-	-	1	-	-	-	-
Tocilizumab	Yokota et al., 2008[29]	Drug	20	1	1	-	-	2	-	-
		Placebo	23	1	1	-	-	4	-	-
Non-biologic agents										
Azathioprine	Kvien et al., 1986[18]	Drug	17	3	-	1	2	-	-	-
		Placebo	15	0	-	-	-	-	-	-
Hydroxychloroquine	Brewer et al., 1986[11]	Drug	57	3	-	2	7	-	-	-
		Placebo	51	3	-	-	4	-	-	-
Methotrexate	Giannini et al., 1992[13]	Drug	86	3	10	0	0	-	-	-
		Placebo	41	0	5	-	-	-	-	-
Penicillamine	Brewer et al., 1986[11]	Drug	54	2	-	4	2	-	-	1
		Placebo	51	2	-	-	4	-	-	-
	Prieur et al., 1985[19]	Drug	38	-	6	3	-	-	-	-
		Placebo	36	-	4	1	-	-	-	-
Sulfasalazine	van Rossum et al., 1998[15]	Drug	35	10	24	9	-	-	9	-
		Placebo	34	0	18	3	-	-	5	-

Table 7 – Part 2. Adverse events reported in RCTs – specific symptoms

DMARD	Study	Intervention	Sample size	Fever	Nausea/vomiting	Pain	Alopecia/Hirsutism	Bleeding	Infection
Biologic agents									
Abatacept	Ruperto et al., 2008[24]	Drug	62	-	-	-	-	-	-
		Placebo	60	-	-	-	-	1	1
Anakinra	Ilowite et al., 2009[30]	Drug	25	3	-	0	-	-	-
		Placebo	25	2	-	2	-	-	-
Etanercept	Lovell et al., 2000[26]	Drug	25	-	-	-	-	-	-
		Placebo	26	-	-	-	-	-	-
	Smith et al., 2005[16]	Drug	7	-	-	-	-	-	5
		Placebo	5	-	-	-	-	-	3
Infliximab + MTX	Ruperto et al., 2007[17]	Drug + MTX	60	-	-	-	-	-	41
		Placebo + MTX	62	-	-	-	-	-	28
IVIG	Giannini et al., 1996[28]	Drug	10	-	-	-	-	-	-
		Placebo	9	-	-	-	-	-	-
	Silverman et al., 1994[20]	Drug	14	-	-	-	-	-	-
		Placebo	17	-	-	-	-	-	-
Tocilizumab	Yokota et al., 2008[29]	Drug	20	-	-	-	-	-	1
		Placebo	23	-	-	-	-	-	1
Non-biologic agents									
Azathioprine	Kvien et al., 1986[18]	Drug	17	1	1	1	1	1	3
		Placebo	15	-	-	2	1	0	-
Hydroxychloroquine	Brewer et al., 1986[11]	Drug	57	-	-	-	-	-	-
		Placebo	51	-	-	-	-	-	-
Methotrexate	Giannini et al., 1992[13]	Drug	86	-	-	6	-	-	-
		Placebo	41	-	-	0	-	-	-
Penicillamine	Brewer et al., 1986[11]	Drug	54	-	-	-	-	-	-
		Placebo	51	-	-	-	-	-	1
	Prieur et al., 1985[19]	Drug	38	-	-	-	-	-	2
		Placebo	36	-	-	-	-	-	-
Sulfasalazine	van Rossum et al., 1998[15]	Drug	35	-	-	-	-	-	-
		Placebo	34	-	-	-	-	-	-

Table 7 – Part 3. Adverse events reported in RCTs – other

DMARD	Study	Intervention	Sample size	Anemia	Other hematologic abnormality	Macrophage activation syndrome	Other laboratory abnormality	Elevated liver enzymes	Other	Serious AEs	Death
Biologic agents											
Abatacept	Ruperto et al., 2008[24]	Drug	62	-	-	-	-	-	-	0	-
		Placebo	60	-	-	-	-	-	-	2	-
Anakinra	Ilowite et al., 2009[30]	Drug	25	-	-	-	-	-	7	-	-
		Placebo	25	-	-	-	-	-	9	-	-
Etanercept	Lovell et al., 2000[26]	Drug	25	-	-	-	-	-	-	-	-
		Placebo	26	-	-	-	-	-	1	-	-
	Smith et al., 2005[16]	Drug	7	-	-	-	-	-	-	-	-
		Placebo	5	-	-	-	-	-	-	-	-
Infliximab + MTX	Ruperto et al., 2007[17]	Drug + MTX	60	-	-	-	-	-	-	19	-
		Placebo + MTX	62	-	-	-	-	-	-	3	1
IVIG	Giannini et al., 1996[28]	Drug	10	-	-	-	-	-	-	-	-
		Placebo	9	-	-	-	-	-	-	-	-
	Silverman et al., 1994[20]	Drug	14	-	-	1	-	1	-	-	-
		Placebo	17	-	-	-	-	-	-	-	-
Tocilizumab	Yokota et al., 2008[29]	Drug	20	-	-	-	-	-	-	-	0
		Placebo	23	-	-	-	-	-	-	-	0
Non-biologic agents											
Azathioprine	Kvien et al., 1986[18]	Drug	17	-	2	-	-	-	1	-	-
		Placebo	15	-	-	-	-	-	0	-	-
Hydroxychloroquine	Brewer et al., 1986[11]	Drug	57	6	4	-	8	-	-	-	-
		Placebo	51	2	2	-	5	-	-	-	-
Methotrexate	Giannini et al., 1992[13]	Drug	86	-	-	-	30	0	3	-	-
		Placebo	41	-	-	-	5	-	0	-	-
Penicillamine	Brewer et al., 1986[11]	Drug	54	2	4	-	9	-	-	-	-
		Placebo	51	2	2	-	5	-	-	-	-
	Prieur et al., 1985[19]	Drug	38	-	1	-	-	-	1	-	-
		Placebo	36	-	-	-	-	-	0	-	-
Sulfasalazine	van Rossum et al., 1998[15]	Drug	35	-	2	-	4	2	-	1	-
		Placebo	34	-	0	-	0	0	-	0	-

Abbreviations to Table 7, Parts 1-3: AEs = adverse events; DMARD = disease-modifying anti-rheumatic drug; IVIG = intravenous immunoglobulin; MTX = methotrexate; RCTs = randomized controlled trials

Placebo-controlled RCTs of Biologic DMARDs

Abatacept

One good-quality study[24] randomized 62 patients to abatacept in a 6-month RCT that was preceded by an open-label run-in phase. No adverse events associated with abatacept were reported.

Anakinra

One study rated as being of fair quality for the purposes of evaluating safety randomized 25 patients to anakinra in a 16-week RCT that was preceded by an open-label run-in phase.[30] Among the patients in the anakinra arm, 6 (24 percent) had gastrointestinal events, 2 (8 percent) had dermatologic events, 8 (32 percent) had respiratory events, 6 (24 percent) had neurologic events, 3 (12 percent) had fever, 2 (6 percent) reported pain, and 7 (42 percent) had other adverse events. None of the adverse events was considered by the authors to be serious. These rates were similar to those observed in the placebo arm, with the exception of the 10 patients (40 percent) who reported dermatologic events.

Etanercept

Two studies compared etanercept to placebo. One[26] was a good-quality study that evaluated only children with polyarticular JRA. Of the 25 patients randomized to the etanercept arm after an open-label run-in phase, gastrointestinal and dermatologic events were each reported in one patient (four percent). There were no dropouts due to adverse events. The second study[16] was a fair-quality RCT that evaluated the safety and effectiveness of etanercept for the treatment of uveitis. Unspecified infections were reported in 5 of the 7 patients (71 percent) in the etanercept arm, and in 3 of the 5 patients (60 percent) in the placebo arm

Infliximab

Infliximab plus methotrexate was compared to placebo plus methotrexate in one fair-quality RCT.[17] This study inconsistently and incompletely reported outcomes, and there was insufficient information to compare adverse event rates in the two study arms over all time periods. Infection was reported in 41 of the 60 patients (68 percent) who received infliximab 3 mg/kg plus methotrexate during the 14 weeks of the RCT phase and the subsequent 38 weeks of the open-label continuation phase, compared to 28 of 62 patients (45 percent) in the placebo plus methotrexate arm during the 14-week RCT phase. Nineteen serious adverse events were reported among the 60 patients (32 percent) in the infliximab plus methotrexate arm over 52 weeks, compared to 3 of 62 patients (5 percent) in the placebo plus methotrexate group over 14 weeks. The nature of the serious adverse events was not reported. Two patients (three percent) in the infliximab plus methotrexate arm and one patient (two percent) in the placebo plus methotrexate arm dropped out because of adverse events.

IVIG

Two studies compared IVIG to placebo. One small, fair-quality study[28] reported no adverse events during the course of the 4-month RCT phase preceded by a 3- to 6-month run-phase among the 10 patients randomized to IVIG. Another study,[20] rated poor in quality, reported macrophage activation syndrome in 1 patient (7 percent) and elevated liver enzymes in another (7 percent) among the 14 patients randomized to IVIG.

Tocilizumab

One fair-quality study compared tocilizumab to placebo during a 12-week double-blind RCT phase preceded by a 6-week run-in phase.[29] One patient in each group (5 percent) dropped out because of adverse events. Of the 20 patients in the tocilizumab arm, 1 (5 percent) reported a gastrointestinal event, 2 (10 percent) reported a respiratory event, and 1 (5 percent) reported a mononucleosis infection.

Placebo-controlled RCTs of Non-Biologic DMARDs

Azathioprine

One fair-quality study compared azathioprine to placebo in a 16-week RCT.[18] Among the 17 patients randomized to azathioprine, 3 (18 percent) dropped out because of adverse events, 3 (18 percent) had an infection, 2 (12 percent) had renal or urologic events, and 2 (12 percent) had a hematologic abnormality. The adverse event rate for dermatologic events, fever, nausea/vomiting, pain, alopecia, or bleeding was 6 percent among patients in the azathioprine arm.

Hydroxychloroquine

One fair-quality RCT compared both hydroxychloroquine and penicillamine to placebo over the course of 12 months.[11] Of the 57 patients in the hydroxychloroquine arm, 3 (5 percent) dropped out due to adverse events, 2 (4 percent) had a dermatologic event, 7 (12 percent) had a renal or urologic event, 6 (11 percent) had anemia, 4 (7 percent) had a hematologic abnormality, and 8 (14 percent) had other laboratory abnormalities. Adverse event rates were similar among patients in the placebo arm.

Methotrexate

A single good-quality study compared methotrexate to placebo in a double-blind RCT of 6 months' duration.[13] Forty-six patients were randomized to low-dose (10 mg/m^2/week) methotrexate, 40 were randomized to very low-dose (5 mg/m^2/week) methotrexate, and 41 were randomized to placebo. Of the 86 patients in a methotrexate arm, 3 (3 percent) dropped out due to adverse events, 10 (12 percent) reported a gastrointestinal event, 6 (7 percent) reported pain, and 30 (35 percent) had a laboratory abnormality (compared to 13 percent in the placebo arm).

Penicillamine

One fair-quality RCT compared both penicillamine and hydroxychloroquine to placebo over the course of 12 months.[11] Of the 51 patients in the penicillamine arm, 2 (4 percent) dropped out do to adverse events, 4 (8 percent) had a dermatologic event, 1 (2 percent) had an ophthalmologic event, 2 (4 percent) had anemia, 4 (8 percent) had a hematologic abnormality, and 9 (17 percent) had other laboratory abnormalities. In another study, a good-quality RCT of 6 months' duration,[19] 38 patients were randomized to the penicillamine arm. Among those patients, 6 (16 percent) reported a gastrointestinal event, 3 (8 percent) reported a dermatologic event, 2 (5 percent) had an infection, and 1 (3 percent) had a hematologic abnormality. Adverse event rates were similar among the patients in the placebo arms in both studies.

Sulfasalazine

A single good-quality RCT of 6 months' duration compared sulfasalazine to placebo.[15] Among the 35 patients randomized to sulfasalazine, 10 (29 percent) dropped out due to adverse events (compared to none in the placebo arm), 24 (69 percent) reported a gastrointestinal event, 9 (26 percent) reported a dermatologic event, 9 (26 percent) reported a neurologic event, 2 (6 percent) had hematologic abnormalities, 2 (6 percent) had elevated liver enzymes, and 4 (11 percent) had other laboratory abnormalities. All of the adverse event rates were higher in the sulfasalazine group than in the placebo group.

Results of Horizon Scan

We performed a horizon scan to identify all published reports of adverse events that may be attributable to DMARDs. The adverse events reported among patients with JIA undergoing treatment with a DMARD are summarized in Appendix E. Patients undergoing treatment with one or more DMARDs in the placebo-controlled RCTs described in the preceding two sections are included in Appendix E. Patients in non-DMARD comparison arms of those RCTs are not, however, included in the Appendix. The "other" category of adverse events includes a wide variety of events that were infrequently reported, such as asthenia, malaise, hostility, or taste disturbance.

A single death possibly associated with DMARD use was reported in a girl on immunosuppressive therapy with cyclosporine A and methotrexate who died of Legionella pneumonia at the age of 53 months.[34] Autopsy revealed stage IV lymphoma that was not previously diagnosed.

An additional eight cases of cancer, five of them lymphomas, were identified: two cases of thyroid carcinoma,[35,36] and one case each of yolk sac carcinoma and non-Hodgkin's lymphoma with etanercept;[36] stage II Hodgkin's lymphoma in a patient who had received both etanercept and methotrexate;[37] Hodgkin's lymphoma with infliximab;[37] and two cases of Hodgkin's lymphoma with methotrexate.[38,39] Of the four cases of Hodgkin's lymphoma, two were associated with methotrexate without a concurrent DMARD, and one occurred in a patient who had been treated with both methotrexate and etanercept.

Apart from than the 8 cases of cancer among the more than 3000 patients represented by the publications identified by our horizon scan, there was no clear evidence of a high incidence or prevalence of any given serious adverse event associated with DMARDs.

Key Question 4. How do the efficacy, effectiveness, safety, and adverse effects of treatment with DMARDs differ among the various categories of JIA?

Key Points

- Insufficient data are available to evaluate the efficacy, effectiveness, safety, or adverse effects of treatment with DMARDs by subtype of JIA.

Detailed Analysis

Literature Identified

The studies considered for this question were those identified for Key Questions 1 and 2, which also included the placebo-controlled trials considered for Key Question 3.

Efficacy and Effectiveness

Only one study compared the effectiveness of the DMARD studied (methotrexate) across different diagnostic subtypes of JIA.[14] There was no statistically significant difference in the effectiveness of methotrexate for oligoarticular JIA versus systemic JIA.

Safety and Adverse Events

The only study we identified that explicitly compared the effectiveness of treatment by diagnostic subtype[14] did not report data on safety data or adverse events. We did not identify any studies that provided reliable information on the comparative safety or rates or types of adverse events among the various subtypes of JIA.

Key Question 5. What is the validity, reliability, responsiveness, and feasibility of the clinical outcomes measures for childhood JIA that are commonly used in clinical trials or within the clinical practice setting?

Key Points:

- The CHAQ was the most extensively evaluated instrument of the priority measures we considered. While it demonstrated high reproducibility and internal consistency, it had only moderate correlations with indices of disease activity and quality of life, and poor to

moderate responsiveness. The CHAQ is sensitive to the degree of disability at baseline, with higher responsiveness for those with initially worse functional impairment.

- In general, reliability was moderate to high for measures of physical function for all measures examined, but poor to moderate for psychosocial domains. Similar findings were noted for measures of validity and responsiveness, where measures of psychosocial function and quality of life showed less correlation with disease activity indices and less responsiveness compared to the physical aspects of JIA. These findings are important to consider when discussing risk and benefits of altering treatments, as patients may have different tradeoffs based on the psychosocial aspects of disease.
- No one instrument or outcome measure appears superior in describing the various aspects of JIA with adequate reliability, validity, and responsiveness.
- Definitions to describe various disease states including improvement, remission, and flare have been developed, but further studies are needed to better define their psychometric properties.

Detailed Analysis

Measures Evaluated

As described in the Methods section, based on our initial review of the literature identified, and in collaboration with the project's technical expert panel (TEP), we selected seven measures for detailed evaluation for Key Question 5. This section provides basic descriptions of these seven measures. While several other outcome instruments have been developed for JIA, including the Juvenile Arthritis Functional Assessment Scale and Report and the Juvenile Arthritis Functionality Scale, their psychometric properties were not independently examined, as they were not selected as priority measures by the TEP.

Measures of Disease Activity

- **Active joint count (AJC):** Standard full joint count assesses 71 possible joints for active disease, defined as joints with swelling or pain/tenderness on range of motion. Limited range of motion may also be assessed, but this is listed as a separate measure from active joint count. This requires a full musculoskeletal exam by a health professional.
- **Physician global assessment of disease activity (PGA):** Typically assessed by asking the physician to rate the child's overall disease activity on a visual analog scale (VAS), with higher scores indicating greater disease activity. Most commonly assessed utilizing a 100 mm VAS; representative anchors are "remission" and "very severe." The same scale is used for all subtypes of JIA.
- **Parent/patient global assessment of well-being (PGW):** Assessed by a VAS, most commonly by asking the parent/caretaker to assess how their child is doing after considering all the ways that arthritis affects their child's life. Representative anchors are "very well" and "very poorly." While the PGA assesses only disease activity, the PGW is an assessment of overall well-being.

Measures of Functional Status/Disability

- **Childhood Health Assessment Questionnaire (CHAQ):** The CHAQ was adapted from the Stanford Health Assessment Questionnaire (HAQ), a validated measure used in adult populations to describe disability quantitatively. The CHAQ focuses on disability and discomfort caused by JIA, which have previously been identified as the major indicators of disease impact. The CHAQ consists of a disability index (CHAQ-DI; 30 items, 8 domains), and two visual analogue scales, one for pain/discomfort (100 mm VAS), and the second for overall well-being (100 mm VAS). The disability index is scored based on the amount of difficulty the child has in completing various tasks. To allow for variation based on the child's age and development, rather than disease status, a "not applicable" category also exists. The instrument is usually completed by parents, although there is a child's form for children over 8 years of age. The CHAQ is scored from 0 to 3, with higher scores indicating greater disability. The CHAQ is widely used and has been validated in multiple languages. A ceiling effect has been noted with the CHAQ, with poor discriminate ability for children with mild functional impairments.

Measures of Health-Related Quality of Life

- **Child Health Questionnaire (CHQ):** The CHQ is a general quality-of-life questionnaire which has been in used in children with JIA. It is a self-administered questionnaire with both a parent form, which is available in two lengths (50 or 28 items) and a child form with 87 items (for children aged > 10 years). Most studies in JIA utilize the 50-item questionnaire for parents. The CHQ addresses multiple domains, including physical functioning, bodily pain or discomfort, general health, range in health, limitations in schoolwork and activities with friends, mental health, behavior, self-esteem, family cohesion, limitations in family activities, and emotional or time impact on parent. Scores range from 0 to 100, with higher scores indicating better well-being. Scores are calculated using equations provided in the CHQ manual. The CHQ is reported as a physical score (CHQ PhS) and a psychosocial score (CHQ PsS), as well as a combined score.
- **Pediatric Quality of Life Inventory (PedsQL) 4.0:** The PedsQL is a self-administered questionnaire consisting of generic core questions and disease-specific questions. It applies to children ages 2 to 18 years and includes both a child and parent component. The generic core has 23 items assessing 4 domains: physical, emotional, social, and school functioning.
- **Pediatric Quality of Life Inventory Rheumatology Module (PedsQL-RM):** The PedsQL-RM consists of 22 items addressing 5 domains: pain and hurt, daily activities, treatment, worry, and communication. The total score is on a 0 to 100 scale, with higher scores indicating better quality of life. The total score is calculated from the physical score and a psychosocial score (average of emotional, social, and school functioning scores).

The above-listed measures are further described and compared in Table 8.

Definitions of Treatment Response now Under Development

In addition to the measures prioritized for detailed evaluation, we identified four developing definitions of treatment response: ACR Pediatric response criteria, a consensus-based definition of remission,[40,41] flare,[42] and minimal disease activity. These definitions are multi-dimensional, often using data from the measures we evaluated in detail.

Table 8. Outcomes measures assessed

Measure/ instrument	Number of items	Domains description	Response categories	Scoring range	Mode of administration	Feasibility	Comments
Measures of disease activity							
Active joint count	Full 76 joints exam	Active arthritis	Active, inactive	0 to 73*	Health professional	Joint count summed	Reduced joint count measures exist
Physician global assessment	1 item	Active disease	Most commonly 100 mm VAS	0 to 100*	Health professional	Measure distance from 0 anchor	
Parent/patient global assessment	1 item	VAS or categorical, overall well-being	Most commonly 100 mm VAS	0 to 100*	Self-administered	Value of VAS, no calculation	Assesses disease activity, functional status, and quality of life
Measures of functional status							
CHAQ	CHAQ-DI: 30 items VAS: - Pain - Overall well-being	Physical function (covering 8 domains) Pain Overall well-being	0 to 3, and NA 0 = no difficulty 3 = inability to perform	Physical function: 0 to 3* VAS: 0-100 mm*	Self-administered, parent or patient	5 minutes to complete Score: highest score in each domain = score for domain; 2 minutes to score	Adapted from Stanford Health Assessment questionnaire
Measures of health-related quality of life							
CHQ	Parent form: 50 or 28 items Child form: 87 items	Physical health Pain Mental health School Social Family	0 to 100 0 = poor well-being 100 = excellent well being	0 to 100^#	Self-administered Children self-administer after age 10 years	Apply scoring formula as per manual	
PedsQL 4.0	23 items	Physical Emotional Social School functioning	5-point Likert scale (never to always)	0 to 100^	Self-administered	Together (generic and rheumatology module) takes 10-15 minutes	
PedsQL-RM	22 items	Pain and hurt Daily activities Treatment Worry Communication	5-point Likert scale (never to always)	0 to 100^	Self-administered		

Abbreviations: CHAQ = Childhood Health Assessment Questionnaire; CHAQ-DI = Childhood Health Assessment Questionnaire Disability Index; CHQ = Child Health Questionnaire; NA = not applicable; PedsQL = Pediatric Quality of Life Inventory; PedsQL-RM = Pediatric Quality of Life Inventory-Rheumatology Module; VAS = visual analog scale

* Higher score equals higher disease activity/functional impairment.
^ Higher score indicates better quality of life.
# Mean score in United States: 50, SD 10.

.

Literature Identified

We identified of 35 publications describing 34 unique studies and involving 14,831 patients that investigated the psychometrics of the selected outcomes measures or developing definitions of treatment response (see Table 9). Among these were 8 studies that evaluated reliability, 20 studies that evaluated validity, and 7 that evaluated responsiveness. Overall, there were 3 RCTs, 12 longitudinal non-randomized trials, 15 cross-sectional studies, and 3 studies with both a longitudinal arm and cross-sectional component. Of our selected outcomes measures, the CHAQ was most extensively studied, with 27 studies. The overall quality of the studies was fair, with few studies commenting on blinding, and only one[43] reporting sample size calculations.

Table 9. Studies of psychometric properties of common JIA outcomes measures and developing definitions of treatment response

Study	Instruments	Psychometrics	N	Study design (followup)	Study population
Outcomes measures of interest					
Bekkering et al., 2007[44]	CHAQ	Reliability Validity	28	Cross-sectional	JIA
Brown et al., 2005[43]	CHAQ	Reliability Responsiveness	92	Longitudinal (6 wk, 6 mo)	JIA
Brunner et al., 2005[45]	CHAQ	Validity	77	Cross-sectional	JRA
Brunner et al., 2005[46]	CHAQ	Responsiveness	92	Longitudinal (3.5 mo)	JRA
Dempster et al., 2001[47]	CHAQ	Reliability Validity Responsiveness	131	Cross-sectional	JRA (spondyloarthropathy)
Geerdink et al., 2009[48]	CHAQ	Validity Feasibility	51	Cross-sectional	JIA
Len et al., 1994[49]	CHAQ	Reliability Validity Feasibility	53	Cross-sectional	JRA
Palmisani et al., 2006[50]	CHAQ	Validity	223	Cross-sectional	JIA
Pouchot et al., 2002[51]	CHAQ	Reliability Validity	306	Cross-sectional	JIA
Pouchot et al., 2004[52]	CHAQ	Validity	306	Cross-sectional	JIA
Saad-Magalhaes et al., 2010[53]	CHAQ	Validity Responsiveness	3193	Mixed cross-sectional and longitudinal (6 mo)	JIA
Singh et al., 1994[54]	CHAQ	Reliability Validity	72	Cross-sectional	JRA
Stephens et al., 2007[55]	CHAQ	Reliability	74	RCT	JIA
Takken et al., 2006[56]	CHAQ	Reliability Validity	76	Mixed cross-sectional and longitudinal	JIA
Tennant et al., 2001[57]	CHAQ	Reliability Validity Responsiveness	53	Cross-sectional	JIA
van der Net et al., 1996[58]	CHAQ	Validity Feasibility	23	Cross-sectional	JCA (polyarthritis)
Cespedes-Cruz et al., 2008[59]	CHQ	Validity Responsiveness	521	RCT (6 mo)	JIA (polyarthritis)
Oliveira et al., 2007[60]	CHQ	Validity	3324	Cross-sectional	JIA
Selvaag et al., 2003[61]	CHQ	Reliability Validity	116	Longitudinal (10 mo)	JRA

Study	Instruments	Psychometrics	N	Study design (followup)	Study population
		Responsiveness			
Sawyer et al., 2005[62]	PedsQL	Reliability Validity	54	Longitudinal (12 mo)	JIA
Bazso et al., 2009[63]	CHAQ , joint count	Validity	434, 3324, 595	Mixed cross-sectional and longitudinal (6 mo)	JIA
Bekkering et al., 2001[64]	CHAQ, joint count	Validity	21	Cross-sectional	JIA (systemic onset JIA)
Magni-Manzoni et al., 2005[65]	CHAQ, joint count, PGA, PGW	Responsiveness	115	Longitudinal	JIA
Ruperto et al., 1999[66]	CHAQ, joint count, PGA, PGW	Reliability Responsiveness	132	Longitudinal (6 mo)	JCA (polyarthritis)
Ruperto et al., 1999[67]	CHAQ, joint count, PGA, PGW	Responsiveness	26	Longitudinal (3 mo)	JCA (oligoarthritis)
Moretti et al., 2005[68]	CHAQ, joint count, PGA, PGW, CHQ	Validity Responsiveness	44	Longitudinal (6 mo)	JIA (oligoarthritis)
Filocamo et al., 2007[69]	CHAQ, PGW, PGA, CHQ	Validity Responsiveness Feasibility	211 [114 longitudinal]	Longitudinal (6 mo)	JIA
Brunner et al., 2004[70]	CHAQ, PedsQL, PGW	Reliability Validity	119	Longitudinal (3.5 mo)	86% JRA
Consolaro et al., 2007[71]	PGA, PGW	Validity	537	Cross-sectional	JIA
Sztajnbok et al., 2007[72]	PGA, PGW	Reliability Validity	197	Cross-sectional	JIA
Developing definitions of treatment response					
Lurati et al., 2006[73]	ACR Pediatric 30, ACR Pediatric 20	Validity	75	Longitudinal	JIA
Giannani et al., 1997[74]	Definition of improvement	Validity	77	Consensus	JRA
Ruperto et al., 1998[75]	Definition of improvement	Validity	111	Longitudinal (6 mo)	JCA (polyarthritis)
Lovell et al., 2000[26] and Brunner et al., 2002[42]	Definition of flare	Validity	25	Randomized discontinuation trial	JRA (polyarthritis)

Abbreviations: ACR Pediatric = American College of Rheumatology Response Criteria; CHAQ = Childhood Health Assessment Questionnaire; CHQ = Child Health Questionnaire; JCA = juvenile chronic arthritis; JIA = juvenile idiopathic arthritis; JRA = juvenile rheumatoid arthritis; mo = month(s); PedsQL = Pediatric Quality of Life Inventory; PGA = physician global assessment of disease activity; PGW = Parent/patient global assessment of well-being; RCT = randomized controlled trial; wk = week(s)

Reliability

Reliability addresses the consistency of the instrument in measuring the construct of interest. We examined three areas of reliability: reproducibility, inter-rater reliability, and internal consistency. Instruments with greater reproducibility and inter-rater reliability may be more feasible to use in clinical trials and require smaller sample sizes to detect clinically important differences between treatment groups. We identified 10 studies examining various aspects of reliability for the CHAQ;[43,44,47,49,51,54-57,70] two studies each for the PGA, PGW[71,72] and PedsQL;[62,70] and one for the CHQ.[61]

Reproducibility, also called test-retest reliability, measures the extent to which an instrument scores the same value on repeat administration, assuming the patient's status is unchanged. This was assessed for the CHAQ in five studies, all of which demonstrated high correlation between administrations (correlation coefficient range 0.79 to 0.96).[44,49,51,54,55] The reliability of the PedsQL and CHQ are less well established in JIA populations. We did not identify any studies reporting reproducibility or internal consistency data in JIA populations for the joint counts, PGA, PGW, CHQ, or PedsQL.

Inter-rater reliability was most commonly explored to determine the correlation between parent and patient scores. Inter-rater reliability was measured for the CHAQ, CHQ, and PedsQL, all of which demonstrated a moderate to strong correlation between parent and child when assessing functional status or disability (CHAQ: 0.54 to 0.84;[43,47,54,70] CHQ PhS: 0.69 to 0.87;[61] PedsQL: 0.46 to 0.8, and PedsQL-RM: 0.3 to 0.90.[62,70] The correlation between parent and child was lower for psychosocial domains in two studies, including the PedsQL-RM worry domain (correlation coefficient 0.3)[62] and the CHQ PsS (correlation coefficient range, 0.38-0.53).[61]

Inter-rater reliability of the global assessment measures (PGA and PGW) was examined through comparisons of the physician and parent assessments, rather than parent/patient. The PGA and PGW were compared in two studies[71,72] and were found to have high rates of discordance. In both studies, discordance was defined as either a parent- or a physician-reported global assessment of 0 (no disease activity/good overall well being) when the other was greater than 0. Based on this definition, both studies demonstrated discordance in 60 percent of participants.

Internal consistency, assessed most commonly using Cronbach's alpha, refers to the extent to which all items measure the same construct. Internal consistency was evaluated in four studies for the CHAQ, with all showing high internal consistency (Cronbach's alpha 0.88 to 0.94 for all domains except the domain for "arising" [0.69]).[51,54,56,57] In addition, shorter versions of the CHAQ-DI were found to have high internal consistency, with Cronbach's alpha of 0.93 for both the 29-item and 18-item instruments.[56]

Validity

Validity refers to how well an instrument measures what it claims to measure. For some outcomes, such as joint inflammation, a reference standard is available (e.g., synovial biopsy) but may not be feasible or acceptable to patients. However, for many of the constructs assessed by the clinical outcome instruments we evaluated, there is no reference standard. Therefore, we

evaluated construct validity based on how well the measures correlated with other indicators of disease, such as global assessments, articular counts, and scores from other validated instruments. We focused on studies in which the psychometric dimensions of the instrument were specifically evaluated for children with JIA. Validation studies looking at the performance of an instrument among rheumatology patients in general, but not specifically in JIA patients, are not included in this review.

Of the 20 articles that met our inclusion criteria, 17 explored validation of the CHAQ[44,45,48-54,56-58,60,63,64,69,72] four validation of the CHQ,[59-61,69] and one validation of the PGA and PGW.[72] In addition, one study focused on the correlation of the PedsQL and PedsQL-RM with pain assessments.[62]

Results are summarized in Table 10. The CHAQ was most strongly correlated with the PGW, with a median correlation of 0.54 (0.44 to 0.7, 6 studies).[45,50,51,53,69,72] Of the articular measures of disease, both the AJC and the joints with limited range of motion (LROM) demonstrated moderate correlations with the CHAQ, with a median correlation of 0.45 (0.14 to 0.67, 9 studies[45,50-54,57,69,72]) and 0.49 (0.3 to0.76, 7 studies[44,45,50,52,60,63,69]), respectively. There was considerable variability in these correlations, with the most significant variations among children categorized by disease duration. For children early in the course of disease, the CHAQ correlated less well with AJC than for children later in the course of disease (0.14 and 0.61, respectively). Those with late disease had a strong correlation with LROM (0.76), but lower correlations with PGA (0.51).[50] Modified forms of the CHAQ, including reduced-item and digital versions, have been validated as well, although the correlation with measures of articular measures is slightly less than for the original CHAQ (values of 0.34 to 0.59).[44,48,56]

While there were no strong correlations between indicators of disease activity and the CHAQ, there were moderately strong correlations with other measures of functional status, including Steinbrocker functional class (Kendall Tau b 0.77).[54] There were also moderate correlations with measures of quality of life, including the PedsQL (-0.62) and the PedsQL-RM (-0.63).[45] Of interest, while there were moderate correlations between the CHAQ and the physical scale of the CHQ (PhS) (-0.58), there was poor correlation with the psychosocial scale of the CHQ (PsS) (-0.25).[61]

Studies of the CHQ reported on the physical scale and psychosocial scales separately. The two studies reporting on validity of the CHQ found consistently higher correlations between the physical component on all measures, from physician and parent/patient global assessments to articular indices and functional status.[60,61] While the CHQ was found to differentiate healthy children from those with JIA, we did not find any results indicating discriminate validity to accurately classify children with JIA by the extent of their disease.[59]

The PedsQL and PedsQL-RM have been studied in the general pediatric rheumatology populations, but the only study focusing on JIA evaluated correlations of both instruments with pain assessments. Child-reported pain assessments correlated with all subscales of the PedsQL and PedsQL-RM, and parent pain assessments correlated with three of four subscales for both instruments.[76]

Table 10. Validity – correlations of instruments with measures of diseases and other instruments

Instrument	PGA median (range)	PGW median (range)	AJC median (range)	LROM median (range)	Swollen joint count median (range)	Other instruments
CHAQ[44,45,48-54,56-58,60,63,64,69,72]	0.45 (0.2 to 0.67) 9 studies	0.54(0.44 to 0.70) 6 studies	0.45 (0.14 to 0.67) 9 studies	0.47 (0.33 to 0.76) 6 studies	0.40 (0.22 to 0.65) 4 studies	PedsQL: -0.62 PedsQL-RM: -0.63 CHQ PhS: -0. 63 and 0.58 (2 studies) CHQ PsS: -0.25 (one study) Steinbrocker functional class: 0.77 Disease Activity Index: 0.60 ACR Functional Class: 0.64 Digital CHAQ: 0.97
CHQ[60,61,69]	CHQ PhS: -0.54 (-0.52 to -0.56) 2 studies CHQ PsS: -0.048 1 study	CHQ PhS: -0.64 (-0.63 to -0.65) 2 studies CHQ PsS: -0.315 1 Study	CHQ PhS: -0.39 (-0.36 to -0.42) 2 studies CHQ PsS: -0.024 1 Study			CHAQ: CHQ PhS: -0.54 (-0.50 to -0.57) 2 studies CHQ PsS: -0.25 (-0.22 to -0.28) 2 studies
PGA[72]			0.47	0.4	0.51	
PGW[72]			0.40	0.38	0.42	

Abbreviations: ACR = American College of Rheumatology; AJC = active joint count; CHAQ = Childhood Health Assessment Questionnaire; CHQ = Child Health Questionnaire; CHQ PhS= Child Health Questionnaire physical score; CHQ PsS = Child Health Questionnaire psychosocial score; LROM = limited range of motion; PedsQL = Pediatric Quality of Life Inventory; PedsQL-RM = Pediatric Quality of Life Inventory Rheumatology Module; PGA = physician global assessment of disease activity; PGW = Parent/patient global assessment of well-being

Responsiveness

Responsiveness is determined by two properties: reproducibility and the ability to register changes in scores when a patient's symptom status shows clinically important improvement or deterioration. Although there is no universally recommended measure of responsiveness, most indices rely on calculation of an effect size. The effect size is a unit-free index that uses the mean change score in the numerator and a measure of variability in the denominator. The standardized response mean (SRM)[77] and the responsiveness index[78,79] are particularly useful approaches to calculating effect sizes for this application because they incorporate information about the response variance into the denominator. According to Cohen and colleagues,[80] an effect size of 0.2 to 0.3 is considered a small effect, around 0.5 (0.4 to 0.7) a medium effect, and 0.8 or above a large effect. Deyo and others argue that the issue is not just sensitivity to change, but the ability to discriminate between those who improve and those who do not.[78,81] Receiver operating characteristic (ROC) curves are proposed as an approach for describing how well various changes in scale scores can distinguish between improved and unimproved patients. This approach requires a valid reference standard to make these clinical classifications.

Responsiveness was assessed in eight studies (Table 11). The responsiveness of the CHAQ was assessed in six studies.[43,53,65,67-69] The results of the six studies were quite variable, with effect sizes ranging from 0 to 0.5. The two studies evaluating responsiveness in oligoarticular populations found the CHAQ was less responsive in patients with oligoarticular disease compared to polyarticular disease, with SRM of 0 to 0.25 for oligoarticular and 0.48 to 0.6 for polyarticular populations.[43,53,65,67,68] This difference in responsiveness by disease subtype was seen even when the same definition of improvement was used.[53,67]

Three studies reported on the responsiveness of the global assessment measures and joint count indices. The most responsive measure was the PGA, with a large effect size, 1.59 (95 percent CI 1.0 to 2.32).[65,67,68] However, in two of these studies, the patients' initial designation as improved or not improved was based on the physician's assessment, either as a categorical assessment on a 5-point scale for the first study,[68] or by a definition of flare based on the addition or escalation of therapy in the second.[65] Swollen joint count and active joint count were also found to have moderate to high responsiveness (effect sizes 1.3 and 0.7, respectively) and may be appropriate alternative measures.[67]

The responsiveness of the CHQ was formally evaluated in two studies, both of which demonstrated poor overall responsiveness, with an SRM of 0.23 and an effect size of 0.18 to 0.23.[61,68] However, in the study that reported responsiveness separately based on disease state, the responsiveness was high in those designated as improved, at 0.96., indicating that the CHQ is sensitive to improvement, but the SRM was lower (-0.60) in those with worsening disease. [61]

The minimum clinically important difference (MCID) was evaluated for the CHAQ in two studies. The MCID helps clinicians interpret study results by estimating the amount of change on an instrument that is associated with a clinically meaningful change in the patient's status. The first study explored the question of minimal clinically important change using a theoretical scenario, and found a mean MCID for improvement of -0.13 in the CHAQ, and 0.75 for worsening.[47] The second study evaluated MCID in a JIA population and found that results differed by which external standard of disease was used, patient, parent, or physician assessment

of disease. The mean MCID for improvement was -0.188 to 0 compared to child ratings, and 0 for parent and physician ratings.[46] The authors concluded that changes in a patient's condition did not correlate well with the CHAQ, and therefore that the CHAQ is unlikely to be to a useful tool when making short-term medical decisions.

The ability of the various outcome measures to differentiate those who improved from those who did not was assessed using ROC curves. In general, ROC curves of 0.5 indicate the measure is no better than chance in discriminating between those who improved compared to those who worsened, while values closer to 1 indicate better discrimination. One study reported on ROC curves for our instruments of interest. The most discriminate measure of the instruments we examined was the physician global assessment, with a ROC curve of 0.86 (95 percent CI 0.72 to 0.95), compared to the parent global assessment value of 0.63 (0.46 to 0.78) and the CHAQ value of 0.56 (0.41 to 0.71).[68]

Table 11. Responsiveness

Instrument	Standardized response means	Effect sizes	ROC curves
CHAQ[43,53,65,67-69]	Median (range): Responders: 0.60 (0.39 to 0.8) Non-responders: 0.08 (0.01 to 0.15)	Median (range): 0.24 (0 to 0.5)	Value (95% CI): 0.56 (0.41 to 0.71)
Physician and Parent/patient global assessments[65,67,68]	Median (range): PGA: 0.9 (0.82 to 2.07) PGW: 0.5 (0.3 to 0.8) Mean change: PGA: 5.4 (2.6) PGW: 1.5 (2.0)	Median (range): PGA: 1.46 (1.0 to 2.32) PGW: 0.5(0.33 to 0.97)	Value (95% CI): PGA: 0.86 (0.72 to 0.95) PGW: 0.63 (0.46 to 0.78)
Joint counts[67]	Number swollen joints: 0.7 Active joints: 1.3	Number swollen joints: 1.3 Active joints: 0.7	
CHQ[61,68]	CHQ PhS: 0.19 CHQ PsS: 0.28 CHQ overall: 0.23	CHQ PhS: 0.18 CHQ PsS: 0.23	CHQ PhS: 0.67(0.5 to 0.81) CHQ PsS: 0.71 (0.54 to 0.85)

Abbreviations: CHAQ = Childhood Health Assessment Questionnaire; CHQ = Child Health Questionnaire; CHQ PhS= Child Health Questionnaire physical score; CHQ PsS = Child Health Questionnaire psychosocial score; CI = confidence interval; PGA = physician global assessment of disease activity; PGW = Parent/patient global assessment of well-being; ROC = receiver operating characteristic

Composite Definitions of Disease Status or Response to Therapy

Because JIA is a complex disorder, several composite definitions have been developed to categorize disease status or response to therapy. We describe these briefly below

American College of Rheumatology Pediatric Response Criteria (ACR Pediatric 30)

The ACR Pediatric 30 response criteria is based on a core set of six variables: (1) physician global assessment of disease activity; (2) parent/patient global assessment of overall well-being; (3) measure of functional ability (CHAQ or JAFAS); (4) number of joints with active arthritis; (5) number of joints with limited range of motion; and (6) ESR. This measure is scored on a relative scale, based on percent improvement or worsening, and was developed to assess

response to therapy in clinical trials. The initial response criteria were developed using a combination of statistical and consensus formation techniques.[74] For each of the 240 definitions of improvement considered, the sensitivity and specificity were calculated using the physicians' consensus rating of improvement as the reference standard. Nine of the definitions with a sensitivity and specificity greater than 80 percent were retained, including the ACR Pediatric 30, which was rated highest based on sensitivity, specificity, measures of agreement, and face validity. The ACR Pediatric 30 is defined as 30 percent or more improvement in three of the six variables, with no more than one variable worsening by more than 30 percent. Similar definitions exist for ACR Pediatric 20, 50, 70, and 90, with the exception of requiring greater percentages of improvement, with no more than one variable worsening by 30 percent or more. These scores provide a relative measure of response, but not current disease state.

Remission

A consensus-based definition of "remission" identifies three subtypes: inactive disease, remission on medications, and remission off medications.[40,41] A Delphi serial questionnaire consensus-formation approach was used to draft the criteria. The criteria for inactive disease include no active arthritis; no fever, rash, splenomegaly, serositis, or generalized lymphadenopathy attributable to JIA; a normal ESR or C-reactive protein; and the best possible score on the physician global assessment of disease activity. In addition, the definition of inactive disease requires there to be no active uveitis. Children with 6 continuous months of inactive disease, as defined above, on medication meet the definition for clinical remission on medication, while 12 months of inactive disease off anti-rheumatic medications defines clinical remission off medication.[40,41] While these definitions have been applied retrospectively to JIA populations, further validations studies are underway.

Flare

A preliminary definition of flare was derived from a cohort of patients with polyarticular JIA using the six core response variables as defined in the ACR Pediatric.[42] The authors defined the standard of flare as treatment with placebo and then examined various definitions of flare based on receiver-operator characteristics. All 25 in the etanercept arm were presumed not to flare; therefore, the specificity of the flare definition equals the number without relapse by the candidate definition divided by the total in the etanercept group. Based on this methodology, a flare was defined as a 40 percent worsening in two of six core set items without improvement in more than one core set variable by 30 percent. This study was based on 51 children, and further validation studies are needed.

Minimal Disease Activity

The authors who defined minimal disease activity (MDA) developed the definition in acknowledgement that many children with JIA do not achieve full remission with current treatments, and that a more reasonable goal for treatment might be minimally active disease.[82] They therefore reviewed patient visits where changes in therapy were initiated verse visits where no change was made or medication was discontinued. They examined measures of disease activity at those visits and established cutoff values that best identified states of MDA. Their

results defined MDA as a physician global assessment of < 2.5 cm and swollen joint count of 0 for oligoarticular disease; and a physician global assessment of < 3.4 cm, parent global assessment < 2.1 cm, and a swollen joint count of < 1 for polyarticular disease.[83] Validation studies are needed.

Summary and Discussion

A succinct summary of the results of this review of the comparative benefits and harms of DMARDs for children with JIA is presented in the tables that follow. First, we provide an aggregated view of the strength of evidence and brief conclusions (Table 12). Next, we describe the nature and quality of the evidence for Key Questions 1, 2, and 4 in a format recommended by the GRADE committee (Tables 13-16). We then provide a tabular summary of the evidence for Key Question 5 (Table 17). Finally, we comment on the applicability of our findings.

Table 12. Summary of the evidence on comparative effectiveness and harms of DMARDs for childhood JIA

Key question	Strength of evidence	Conclusions
1. In children with JIA, does treatment with DMARDs, compared to conventional treatment:		
a. Improve laboratory measures of inflammation?	Low	Trials of DMARDs usually report changes in laboratory measures of inflammation (e.g., ESR). However, ESR is inconsistently associated with treatment. This is based on 14 studies of 1060 subjects.
b. Improve radiological progression?	Insufficient	Insufficient data are available to evaluate the impact of DMARDs on radiological progression. Only one cohort study of 63 subjects reported data on radiological progression.
c. Improve symptoms?	Moderate	Among children who have responded to a biologic DMARD, randomized discontinuation trials show that continued treatment for 4 months to 2 years decrease the risk of having a flare (RR 0.46, 95% CI 0.36 to 0.60). This is based on four studies of 322 subjects. Among the non-biologic DMARDs, there is some evidence that methotrexate is superior to conventional therapy and oral corticosteroids based on two randomized trials of 215 subjects.
d. Improve health status?	Low	Changes in health status were reported in 12 studies involving 927 subjects. Health status improved inconsistently with treatment with DMARDs.
2. In children with JIA, what are the comparative effects of DMARDs on:		
a. Laboratory measures of inflammation?	Low	Trials of DMARDs usually report changes in laboratory measures of inflammation (e.g., ESR). However, ESR is inconsistently associated with treatment. This is based on 4 RCTs of 448 subjects and 1 cohort study of 72 subjects.

Key question	Strength of evidence	Conclusions
b. Radiological progression?	Insufficient	No study addressed radiologic progression.
c. Symptoms?	Low	The non-biologic DMARDs that were compared directly (penicillamine vs. hydroxychloroquine, sulfasalazine vs. hydroxychloroquine, and leflunomide vs. methotrexate) had similar effectiveness. Changes in symptoms between the treatment arms were not measured with significant precision to detect a difference. This is based on 4 RCTs of 448 subjects and 1 cohort study of 72 subjects. One poor-quality RCT of 94 subjects found that etanercept was similar to infliximab.
d. Health status?	Low	The non-biologic DMARDs that were compared directly (penicillamine vs. hydroxychloroquine, sulfasalazine vs. hydroxychloroquine, and leflunomide vs. methotrexate) had similar effectiveness. Changes in health status between the treatment arms were not measured with significant precision to detect a difference. This is based on 4 RCTs of 448 subjects and 1 cohort study of 72 subjects. One poor quality RCT of 94 subjects found that etanercept was similar to infliximab.
3. In children with JIA, does the rate and type of adverse events differ between:		
a. The various DMARDs?	Insufficient	Three RCTs directly compared two DMARDs; two compared penicillamine to hydroxychloroquine, and one compared leflunomide to methotrexate. The rate and type of adverse events did not differ between treatment groups in these studies. High variability across studies in the ascertainment and reporting of adverse events preclude valid comparisons of the rate and type of adverse events among the various DMARDs.
b. DMARDs and conventional treatment with or without methotrexate?	Insufficient	No RCT directly compared a DMARD to conventional treatment. Thirteen trials directly compared a DMARD to placebo. The rate and type of adverse events were generally similar between intervention and placebo groups, with the notable exceptions of infliximab plus methotrexate being associated with more serious adverse events (32% vs. 5% over differing lengths of followup), and methotrexate being associated with higher rates of laboratory abnormalities (35% vs. 13%).
4. How do the efficacy, effectiveness, safety, and adverse effects of treatment with DMARDs differ among the various categories of JIA?	Insufficient	Only one study – an RCT of methotrexate versus placebo in which each group could also receive oral corticosteroids, intra-articular corticosteroids, and NSAIDs – evaluated efficacy by JIA subtype. No difference was found among those with extended oligoarticular JIA (n = 43) and systemic JIA (n = 45). We did not identify any studies that provide reliable

Key question	Strength of evidence	Conclusions
		information on the comparative safety or rates or types of adverse events among the various subtypes of JIA.
5. What is the validity, reliability, responsiveness, and feasibility of the clinical outcome measures for childhood JIA that are commonly used in clinical trials or within the clinical practice setting?	Insufficient	Most of the studies examining the psychometric properties of the instruments used in JIA were fair-quality cross-sectional or longitudinal non-randomized controlled trials. No one instrument or outcomes measure appeared superior in measuring disease activity or functional status. The current response criteria of the ACR Pediatric 30, a composite measure that includes articular indices, functional status, laboratory measure, and global assessments, takes into account the various measures most commonly used. However, the responsiveness of several of these measures, including functional status and parent/patient global assessment, are poor to moderate, and they may not adequately reflect changes in disease state. Furthermore, given that the ACR Pediatric 30 is a relative measure of disease activity, the impact of JIA subtype on percent improvement is unclear, as certain instruments, such as the CHAQ, appear to have differential responsiveness depending on extent of disease at baseline. The ACR Pediatric 30 is also a relative measure of disease activity and not a measure of current disease state.

Abbreviations: ACR = American College of Rheumatology; CHAQ = Childhood Health Assessment Questionnaire; CI = confidence interval; DMARD(s) = disease-modifying anti-rheumatic drug(s); ESR = erythrocyte sedimentation rate; JIA = juvenile idiopathic arthritis; NSAIDs = non-steroidal anti-inflammatory drugs; RCT = randomized controlled trial; RR = risk ratio

GRADE summary tables were developed to describe the strength of evidence. For Key Question 1, separate GRADE summary tables are presented for the biologic and non-biologic DMARDs. We identified six randomized discontinuation trials that were conducted for the biologic DMARDs. Unlike RCTs or prospective cohort trials, randomized discontinuation trials evaluate the risk of worsening disease among those who initially responded to therapy. Because of this fundamental difference, we present a separate GRADE strength of evidence rating for the randomized discontinuation studies for each outcome. GRADE summary tables do not apply to Key Question 3 or Key Question 5. Findings from Key Question 3 are summarized in Tables 7 (Parts 1-3), under Results, and in Appendix E. Findings from Key Question 5 are summarized by outcome measure in Table 16.

Table 13. GRADE summary table for Key Question 1 – biologic DMARDs

	Domains pertaining to strength of evidence					
Number of studies; subjects	**Design**	**Risk of bias/study quality**	**Consistency**	**Directness**	**Precision**	**Strength of evidence**
Laboratory measures of inflammation						Low
1; 31	RCT	Poor (high dropout rate)	NA	Direct	Imprecise	-
4; 322	Randomized discontinuation trials	Poor to Good	Inconsistent	Direct	Precise	
1; 20	Cohort	Poor (open-label)	NA	Direct	Imprecise	-
Radiologic progression						Insufficient
0; 0	RCT	-	-	-	-	-
0; 0	Randomized discontinuation trials	-	-	-	-	-
0;0	Cohort	--	--	--	--	-
Symptoms						Moderate
3; 165	RCT	Fair (one study had high dropout rate)	Consistent	Direct	Imprecise	-
6; 341	Randomized discontinuation trials	Good	Consistent	Direct	Precise	
0; 0	Cohort	-	-	-	-	-
Health status						Low
1; 31	RCT	Poor (high dropout rate)	NA	Direct	Imprecise	-
4; 272	Randomized discontinuation trials	Good	Inconsistent	Direct	Imprecise	
0; 0	Cohort	-	-	-	-	-

Abbreviations: DMARDs = disease-modifying anti-rheumatic drugs; NA = not applicable; RCT = randomized controlled trial

Table 14. GRADE summary table for Key Question 1 – non-biologic DMARDs

	Domains pertaining to strength of evidence					
Number of studies; subjects	**Design**	**Risk of bias/study quality**	**Consistency**	**Directness**	**Precision**	**Strength of evidence**
Laboratory measures of inflammation						Low
7; 624	RCT	Fair (open-label or unblinded)	Inconsistent	Direct	Imprecise	-
1; 63	Cohort	Poor (open-label)	NA	Direct	Imprecise	-
Radiologic progression						Low
1; 69	RCT	Good	NA	Direct	Imprecise	-
0; 0	Cohort	-	-	-	-	-
Symptoms						Moderate (MTX) Low (other non-biologic)
7; 624	RCT	Fair (open-label or unblinded)	Consistent	Direct	Imprecise	-
1; 63	Cohort	Poor (open-label)	NA	Indirect	Imprecise	-
Health status						Moderate (MTX) Low (other non-biologic)
7; 624	RCT	Fair (open-label or unblinded)	Consistent	Direct	Imprecise	-
1; 63	Cohort	Poor (open-label)	NA	Indirect	Imprecise	-

Abbreviations: DMARDs = disease-modifying anti-rheumatic drugs; MTX = methotrexate; NA = not applicable; RCT = randomized controlled trial

Table 15. GRADE summary table for Key Question 2

	Domains pertaining to strength of evidence					
Number of studies; subjects	**Design**	**Risk of bias/study quality**	**Consistency**	**Directness**	**Precision**	**Strength of evidence**
Laboratory measures of inflammation						Low
4; 448	RCT	Fair (some studies with incomplete blinding)	Inconsistent	Direct	Imprecise	-
1; 72	Cohort	Poor (insufficient data)	NA	Direct	Imprecise	-
Radiologic progression						Insufficient
0; 0	RCT	-	-	-	-	-
0; 0	Cohort	-	-	-	-	-
Symptoms						Low
4; 448	RCT	Fair (some studies with incomplete blinding)	Inconsistent	Direct	Imprecise	-
1; 72	Cohort	Poor (insufficient data)	NA	Direct	Imprecise	-
Health status						Low
4; 448	RCT	Fair (some studies with incomplete blinding)	Inconsistent	Direct	Imprecise	-
1; 72	Cohort	Poor (insufficient data)	NA	Direct	Imprecise	-

Abbreviations: NA = not applicable; RCT = randomized controlled trial

Table 16. GRADE summary table for Key Question 4

	Domains pertaining to strength of evidence					
Number of studies; subjects	**Design**	**Risk of bias/study quality**	**Consistency**	**Directness**	**Precision**	**Strength of evidence**
Laboratory measures of inflammation						Insufficient
1; 88	RCT	Good	NA	Direct	Imprecise	-
0; 0	Cohort	-	-	-	-	-
Radiologic progression						Insufficient
0; 0	RCT	-	-	-	-	-
0; 0	Cohort	-	-	-	-	-
Symptoms						Insufficient
1; 88	RCT	Good	NA	Direct	Imprecise	-
0; 0	Cohort	-	-	-	-	-
Health status						Insufficient
1; 88	RCT	Good	NA	Direct	Imprecise	-
0; 0	Cohort	-	-	-	-	-

Abbreviations: NA = not applicable; RCT = randomized controlled trial

Table 17. Evidence summary table for Key Question 5

Number of studies; subjects	Evidence summary
Active joint count	
12; 8064	Shows high responsiveness and moderate correlation with other measures of disease activity and functional status, but poor correlation with psychosocial aspects of quality of life. Lack of inter-rater reliability data.
Physician global assessment of disease activity	
12; 8668	Moderate correlations with measures of disease activity, the CHAQ, and quality-of-life measures. Responsiveness difficult to measure, as often compared to other physician measures of disease activity. No data on inter-rater reliability between providers.
Parent/patient global assessment of well-being	
8; 8182	Moderate correlations with other measures of disease activity, the CHAQ, and physical aspects of the quality of life measures, but poor correlation with psychosocial aspects of the CHQ. Moderate responsiveness and discriminate abilities.
CHAQ	
25; 13374	Most commonly reported outcome measure with strong reliability, including moderate to strong inter-rater reliability between parent and child. Moderate correlations with other measures of disease activity, but poor responsiveness, which varies depending on how extensive the arthritis is at baseline (ceiling effect).
CHQ	
5; 4687	Limited data for JIA population. Moderate to strong parent to child inter-rater reliability for physical components, but lower for psychosocial aspects. Similarly, moderate correlations with measures disease activity, and the CHAQ for the physical component of the CHQ, but poor for the psychosocial domains. Poor responsiveness.
PedsQL/PedsQL-RM	
2; 173	Insufficient data in JIA populations to evaluate fully. Moderate to strong parent to child inter-rater reliability for physical components, but lower for psychosocial aspects.

Abbreviations: CHAQ = Childhood Health Assessment Questionnaire; CHQ = Child Health Questionnaire; JIA = juvenile idiopathic arthritis; PedsQL = Pediatric Quality of Life Inventory; PedsQL-RM = Pediatric Quality of Life Inventory-Rheumatology Module

Applicability was assessed for Key Question 1 only. Insufficient evidence was available to rate applicability for Key Questions 2 and 3, and Key Question 4 and 5 were not amenable to assessment of applicability. For Key Question 1, we assessed applicability as follows:

- Population: There was variation across studies in the definition of JIA and both duration and severity of illness, likely reflecting the range of patients seen in usual practice. However, six of the studies of the biologic DMARDs were randomized discontinuation studies, which include patients who have responded to the intervention.
- Intensity or quality of treatment: With the exception of methotrexate, the non-biologic DMARDs are less often used than the newer biologic DMARDs. The intensity of treatment in the studies of the biologic DMARDs is consistent with current recommendations.
- Choice of, and dosing of, the comparator: Methotrexate, a non-biologic DMARD, is a standard treatment for JIA. Six of the studies of the biologic DMARDs included methotrexate as a comparator; none of the studies of the non-biologic DMARDs included methotrexate as a comparator. The reasons for use or dose escalation of the comparator drugs were usually not described.
- Outcomes: The most commonly reported outcome measures were laboratory indicators of inflammation (e.g., ESR) or the ACR Pediatric 30. The ACR Pediatric 30 blends several relevant outcomes (e.g., active joint count, functional status, pain), but is not normally used in daily clinical practice. As described for Key Question 5, new instruments to better assess response to therapy and changes in health-related quality of life are in development.
- Timing of followup: Five of the studies of biologic DMARDs and five of the studies of non-biologic DMARDs actively followed subjects for more than 6 months. This would allow sufficient time to detect clinically important benefits and may be long enough to identify important harms.

Future Research

Efficacy of DMARDs

Although DMARDs have improved health outcomes for children with JIA, few data are available to evaluate the comparative effectiveness of either specific DMARDs or general classes of DMARDs (e.g., non-biologic vs. biologic, or by mechanism of action). Research on the effectiveness of treatments for JIA is challenging because it is a rare condition that includes multiple subtypes, which could potentially respond differentially to therapy. Furthermore, the health impact of JIA fluctuates over time. Therefore, trials require large sample sizes with long follow-up periods.

Developing a summary estimate of effectiveness of the DMARDs is challenging because there is:

- Heterogeneity in the study population. Changes in the definition of JIA (e.g., JRA, JCA) may have led to the inclusion in studies of individuals who may respond differently to treatments. Similarly, differences by disease subtype (e.g., polyarticular, pauciarticular, systemic) might lead to different conclusions about the effectiveness of treatment.
- Variation in comparators. Over time, the standard of care for JIA has changed. For example, relatively recent studies of biologic DMARDs often allow methotrexate, a DMARD, in the comparator group, while older studies do not include methotrexate in the comparator groups. Some older studies included systemic corticosteroids as a comparator.
- Outcome measures vary across the studies and are sometimes incompletely described. For example, some studies report the percentage improvement from baseline without providing baseline data or an estimate of variability. Among the six randomized discontinuation trials, for example, four reported laboratory measures of inflammation,[24,26,29,30] four reported whether a flare occurred,[24-26,30] three reported active joint count,[24,28,29] and four reported quality of life as measured by CHAQ.[24,26,29,30] Of those that reported the CHAQ score, one[26] reported only the percentage change from baseline without the absolute value or measure of dispersion (e.g., range, standard deviation), and two[29,30] gave average values without measures of dispersion.

Future trials in this domain should consider:

- The challenge of the appropriate comparator. Trials are needed to evaluate the effectiveness of DMARDs compared to conventional therapy as well as against other DMARDs. Defining conventional therapy is challenging because it evolves with advances in the field. Factorial designs involving multiple treatments are a potential solution. Patient-level meta-analysis, pre-planned across different trials, may also help address this issue.
- The issue of treatment-by-subtype interaction. To fully explore comparative effectiveness, larger studies will be needed. In addition, patient-level meta-analysis may help address this challenge.
- The need for study populations who are representative of typical patients with JIA. Subjects from the studies included in this review were identified through specialty

clinics, which is appropriate for rare conditions. However, baseline characteristics varied. Studies should be designed to reflect the comparative effectiveness for typical subjects at various points along the disease spectrum (e.g., at presentation, after failing conventional treatment).

- The variable course of JIA. Trials that evaluate the efficacy of treatment should be sufficiently long with frequent assessment of health status to capture the natural variability of the disease course.
- The need for standardized definitions for, and systematic ascertainment and reporting of, adverse events possibly associated with therapeutic interventions in the treatment of JIA.
- The impact of DMARDs on the specific health conditions associated with JIA, including uveitis and macrocyte activation syndrome.

Study designs other than RCTs will be important in understanding the role of DMARDs in JIA. Randomized discontinuation trials have helped to define the risk of flare in patients who respond to a particular DMARD. Large cohort studies will be important for evaluating the risk of adverse events associated with DMARDs. Such studies could also be important for better characterizing long-term outcomes in JIA.

Safety of DMARDs

Few high-quality data are available regarding the adverse events associated with DMARDs. Because JIA is a chronic illness, understanding the long-term effects of these drugs is critical. One solution to evaluating risk would be to develop registries for DMARDs when used for childhood JIA. Understanding such risk will also provide information about the sequence in which these drugs should be used for difficult-to-treat JIA, or the impact of using multiple drugs.

Our findings suggest that short-term mortality rates associated with DMARDs are very low (we identified only a single patient among more than 3000 who died shortly after receiving a DMARD). The incidence of malignancies during a short course of DMARD treatment also appears to be very low. However, the available evidence is inadequate to determine whether the rates and types of adverse events differ between the various DMARDs or between DMARDs and conventional treatment. The findings from RCTs do not reveal a clear pattern pertaining to adverse events associated with the treatment of JIA with DMARDs compared to placebo. The horizon scan revealed marked differences in the rate and type of adverse event by DMARD, but these findings should be interpreted with caution for several reasons, including: variable definitions of adverse events across studies; non-systematic methods of ascertaining adverse events; nearly universal lack of standard reporting of serious adverse events; a predominance of case reports and uncontrolled series; small sample sizes in most series and RCTs; a limited number of studies for many individual DMARDs; and frequent use of multiple medications and other co-interventions.

Impact of DMARDs on Health Outcomes

Our findings suggest the need for better clinical outcomes measures that are responsive to change across the full spectrum of disease severity. Consistent use of such outcomes measures would facilitate comparative effectiveness research.

The heterogeneity in disease severity and the broad impact of the disease on both physical and psychosocial aspects of children's lives make it difficult to accurately assess children using one instrument or measure. Given the complex nature of JIA, with the potential for both chronic and acute functional limitations and pain, it is difficult to find one tool or instrument that can be responsive to all the facets of disease. Efforts to develop a more standardized composite measure which could incorporate articular indices, severity, and a broader assessment of functional limitations and psychosocial impact would be useful to better differentiate levels of disease activity and overall impact of disease. The current response criteria of the ACR Pediatric 30 definition of improvement, a composite measure which includes articular indices, functional status, laboratory measure, and global assessments, takes into account the various measures most commonly used. However, the responsiveness of several of these measures, including functional status and parent/patient global assessment, are poor to moderate, and they may not adequately reflect changes in disease state. Furthermore, given that the ACR Pediatric 30 is a relative measure of disease activity, the impact of JIA subtype on percent improvement is unclear, as certain instruments, such as the CHAQ, appear to have differential responsiveness by extent of disease at baseline. The ACR Pediatric 30 is also a relative measure and not a measure of current disease state.

Developing an instrument or composite measure to accurately describe all the aspects of JIA, including disease activity, functional status, and quality of life would improve our understanding of the overall impact of JIA. In addition, focusing on the most responsive outcome measures to assess treatment effects would enhance our ability to detect promising new treatments.

References

1. Manners PJ, Bower C, Manners PJ, et al. Worldwide prevalence of juvenile arthritis why does it vary so much? J Rheumatol 2002;29(7):1520-30.
2. Helmick CG, Felson DT, Lawrence RC, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis Rheum 2008;58(1):15-25.
3. Karvonen M, Viik-Kajander M, Moltchanova E, et al. Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. Diabetes Care 2000;23(10):1516-26.
4. Goldmuntz EA, White PH, Goldmuntz EA, et al. Juvenile idiopathic arthritis: a review for the pediatrician. Pediatr Rev 2006;27(4):e24-32.
5. Agency for Healthcare Research and Quality. Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews, Version 1.0 [Draft posted Oct. 2007]. Rockville, MD: Agency for Healthcare Research and Quality. Available at: http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf. Accessed July 29, 2010.
6. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Medical Research Methodology 2006;6:9.
7. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328(7454):1490.
8. Anonymous. Review Manager (RevMan) [Computer program]. Version 5.0. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008.
9. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-58.
10. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-60.
11. Brewer EJ, Giannini EH, Kuzmina N, et al. Penicillamine and hydroxychloroquine in the treatment of severe juvenile rheumatoid arthritis. Results of the U.S.A.-U.S.S.R. double-blind placebo-controlled trial. N Engl J Med 1986;314(20):1269-76.
12. van Kerckhove C, Giannini EH, Lovell DJ. Temporal patterns of response to D-penicillamine, hydroxychloroquine, and placebo in juvenile rheumatoid arthritis patients. Arthritis Rheum 1988;31(10):1252-8.
13. Giannini EH, Brewer EJ, Kuzmina N, et al. Methotrexate in resistant juvenile rheumatoid arthritis. Results of the U.S.A.-U.S.S.R. double-blind, placebo-controlled trial. The Pediatric Rheumatology Collaborative Study Group and The Cooperative Children's Study Group. N Engl J Med 1992;326(16):1043-9.
14. Woo P, Southwood TR, Prieur AM, et al. Randomized, placebo-controlled, crossover trial of low-dose oral methotrexate in children with extended oligoarticular or systemic arthritis. Arthritis Rheum 2000;43(8):1849-57.
15. van Rossum MA, Fiselier TJ, Franssen MJ, et al. Sulfasalazine in the treatment of juvenile chronic arthritis: a randomized, double-blind, placebo-controlled, multicenter study. Dutch Juvenile Chronic Arthritis Study Group. Arthritis Rheum 1998;41(5):808-16.
16. Smith JA, Thompson DJ, Whitcup SM, et al. A randomized, placebo-controlled, double-masked clinical trial of etanercept for the treatment of uveitis associated with juvenile idiopathic arthritis. Arthritis Rheum 2005;53(1):18-23.
17. Ruperto N, Lovell DJ, Cuttica R, et al. A randomized, placebo-controlled trial of infliximab plus methotrexate for the treatment of polyarticular-course juvenile rheumatoid arthritis. Arthritis Rheum 2007;56(9):3096-106.
18. Kvien TK, Hoyeraal HM, Sandstad B. Azathioprine versus placebo in patients with juvenile rheumatoid arthritis: a single center double blind comparative study. J Rheumatol 1986;13(1):118-23.
19. Prieur AM, Piussan C, Manigne P, et al. Evaluation of D-penicillamine in juvenile chronic arthritis. A double-blind, multicenter study. Arthritis Rheum 1985;28(4):376-82.
20. Silverman ED, Cawkwell GD, Lovell DJ, et al. Intravenous immunoglobulin in the treatment of systemic juvenile rheumatoid arthritis: a randomized placebo controlled trial. Pediatric Rheumatology Collaborative Study Group. J Rheumatol 1994;21(12):2353-8.
21. Kvien TK, Hoyeraal HM, Sandstad B. Slow acting antirheumatic drugs in patients with juvenile rheumatoid arthritis--evaluated in a randomized, parallel 50-week clinical trial. J Rheumatol 1985;12(3):533-9.
22. Oppermann J, Mobius D. Therapeutical and immunological effects of methylprednisolone pulse therapy in comparison with intravenous immunoglobulin. Treatment in patients with juvenile chronic arthritis. Acta Univ Carol Med (Praha) 1994;40(1-4):117-21.
23. Riddle R, Ryser CN, Morton AA, et al. The impact on health-related quality of life from non-steroidal anti-inflammatory drugs, methotrexate, or steroids in treatment for juvenile idiopathic arthritis. J Pediatr Psychol 2006;31(3):262-71.
24. Ruperto N, Lovell DJ, Quartier P, et al. Abatacept in children with juvenile idiopathic

arthritis: a randomised, double-blind, placebo-controlled withdrawal trial. Lancet 2008;372(9636):383-91.
25. Lovell DJ, Ruperto N, Goodman S, et al. Adalimumab with or without methotrexate in juvenile rheumatoid arthritis. N Engl J Med 2008;359(8):810-20.
26. Lovell DJ, Giannini EH, Reiff A, et al. Etanercept in children with polyarticular juvenile rheumatoid arthritis. Pediatric Rheumatology Collaborative Study Group. N Engl J Med 2000;342(11):763-9.
27. Lovell DJ, Giannini EH, Reiff A, et al. Long-term efficacy and safety of etanercept in children with polyarticular-course juvenile rheumatoid arthritis: interim results from an ongoing multicenter, open-label, extended-treatment trial. Arthritis Rheum 2003;48(1):218-26.
28. Giannini EH, Lovell DJ, Silverman ED, et al. Intravenous immunoglobulin in the treatment of polyarticular juvenile rheumatoid arthritis: a phase I/II study. Pediatric Rheumatology Collaborative Study Group. J Rheumatol 1996;23(5):919-24.
29. Yokota S, Imagawa T, Mori M, et al. Efficacy and safety of tocilizumab in patients with systemic-onset juvenile idiopathic arthritis: a randomised, double-blind, placebo-controlled, withdrawal phase III trial. Lancet 2008;371(9617):998-1006.
30. Ilowite N, Porras O, Reiff A, et al. Anakinra in the treatment of polyarticular-course juvenile rheumatoid arthritis: safety and preliminary efficacy results of a randomized multicenter study. Clin Rheumatol 2009;28(2):129-37.
31. Lahdenne P, Vahasalo P, Honkanen V. Infliximab or etanercept in the treatment of children with refractory juvenile idiopathic arthritis: an open label study. Ann Rheum Dis 2003;62(3):245-7.
32. Hoza J, Kadlecova T, Nemcova D, et al. Sulphasalazine and Delagil--a comparative study in patients with juvenile chronic arthritis. Acta Univ Carol Med (Praha) 1991;37(1-2):80-3.
33. Silverman E, Mouy R, Spiegel L, et al. Leflunomide or methotrexate for juvenile rheumatoid arthritis. N Engl J Med 2005;352(16):1655-66.
34. Krugmann J, Sailer-Hock M, Muller T, et al. Epstein-Barr virus-associated Hodgkin's lymphoma and legionella pneumophila infection complicating treatment of juvenile rheumatoid arthritis with methotrexate and cyclosporine A. Hum Pathol 2000;31(2):253-5.
35. Horneff G, Schmeling H, Biedermann T, et al. The German etanercept registry for treatment of juvenile idiopathic arthritis. Ann Rheum Dis 2004;63(12):1638-44.
36. Horneff G, De Bock F, Foeldvari I, et al. Safety and efficacy of combination of etanercept and methotrexate compared to treatment with etanercept only in patients with juvenile idiopathic arthritis (JIA): preliminary data from the German JIA Registry. Ann Rheum Dis 2009;68(4):519-25.
37. Yildirim-Toruner C, Kimura Y, Rabinovich E. Hodgkin's lymphoma and tumor necrosis factor inhibitors in juvenile idiopathic arthritis. J Rheumatol 2008;35(8):1680-1.
38. Cleary AG, McDowell H, Sills JA. Polyarticular juvenile idiopathic arthritis treated with methotrexate complicated by the development of non-Hodgkin's lymphoma. Arch Dis Child 2002;86(1):47-9.
39. Takeyama J, Sato A, Nakano K, et al. Epstein-Barr virus associated Hodgkin lymphoma in a 9-year-old girl receiving long-term methotrexate therapy for juvenile idiopathic arthritis. J Pediatr Hematol Oncol 2006;28(9):622-4.
40. Wallace CA, Ruperto N, Giannini E, et al. Preliminary criteria for clinical remission for select categories of juvenile idiopathic arthritis. J Rheumatol 2004;31(11):2290-4.
41. Wallace CA, Ravelli A, Huang B, et al. Preliminary validation of clinical remission criteria using the OMERACT filter for select categories of juvenile idiopathic arthritis. J Rheumatol 2006;33(4):789-95.
42. Brunner HI, Lovell DJ, Finck BK, et al. Preliminary definition of disease flare in juvenile rheumatoid arthritis. J Rheumatol 2002;29(5):1058-64.
43. Brown GT, Wright FV, Lang BA, et al. Clinical responsiveness of self-report functional assessment measures for children with juvenile idiopathic arthritis undergoing intraarticular corticosteroid injections. Arthritis Rheum 2005;53(6):897-904.
44. Bekkering WP, ten Cate R, van Rossum MA, et al. A comparison of the measurement properties of the Juvenile Arthritis Functional Assessment Scale with the childhood health assessment questionnaire in daily practice. Clin Rheumatol 2007;26(11):1903-7.
45. Brunner HI, Johnson AL, Barron AC, et al. Gastrointestinal symptoms and their association with health-related quality of life of children with juvenile rheumatoid arthritis: validation of a gastrointestinal symptom questionnaire. J Clin Rheumatol 2005;11(4):194-204.
46. Brunner HI, Klein-Gitelman MS, Miller MJ, et al. Minimal clinically important differences of the childhood health assessment questionnaire. J Rheumatol 2005;32(1):150-61.
47. Dempster H, Porepa M, Young N, et al. The clinical meaning of functional outcome scores in children with juvenile arthritis. Arthritis Rheum 2001;44(8):1768-74.
48. Geerdink LM, Prince FH, Looman CW, et al. Development of a digital Childhood Health Assessment Questionnaire for systematic monitoring of disease activity in daily practice. Rheumatology (Oxford) 2009;48(8):958-63.

49. Len C, Goldenberg J, Ferraz MB, et al. Crosscultural reliability of the Childhood Health Assessment Questionnaire. J Rheumatol 1994;21(12):2349-52.
50. Palmisani E, Solari N, Magni-Manzoni S, et al. Correlation between juvenile idiopathic arthritis activity and damage measures in early, advanced, and longstanding disease. Arthritis Rheum 2006;55(6):843-9.
51. Pouchot J, Larbre JP, Lemelle I, et al. Validation of the French version of the Childhood Health Assessment Questionnaire (CHAQ) in juvenile idiopathic arthritis. Joint Bone Spine 2002;69(5):468-81.
52. Pouchot J, Ecosse E, Coste J, et al. Validity of the childhood health assessment questionnaire is independent of age in juvenile idiopathic arthritis. Arthritis Rheum 2004;51(4):519-26.
53. Saad-Magalhaes C, Pistorio A, Ravelli A, et al. Does removal of aids/devices and help make a difference in the Childhood Health Assessment Questionnaire disability index? Ann Rheum Dis 2010;69(1):82-7.
54. Singh G, Athreya BH, Fries JF, et al. Measurement of health status in children with juvenile rheumatoid arthritis. Arthritis Rheum 1994;37(12):1761-9.
55. Stephens S, Singh-Grewal D, Bar-Or O, et al. Reliability of exercise testing and functional activity questionnaires in children with juvenile arthritis. Arthritis Rheum 2007;57(8):1446-52.
56. Takken T, van den Eijkhof F, Hoijtink H, et al. Examining the psychometric characteristics of the Dutch childhood health assessment questionnaire: room for improvement? Rheumatol Int 2006;26(11):979-83.
57. Tennant A, Kearns S, Turner F, et al. Measuring the function of children with juvenile arthritis. Rheumatology (Oxford) 2001;40(11):1274-8.
58. van der Net J, Prakken AB, Helders PJ, et al. Correlates of disablement in polyarticular juvenile chronic arthritis--a cross-sectional study. Br J Rheumatol 1996;35(1):91-100.
59. Cespedes-Cruz A, Gutierrez-Suarez R, Pistorio A, et al. Methotrexate improves the health-related quality of life of children with juvenile idiopathic arthritis. Ann Rheum Dis 2008;67(3):309-14.
60. Oliveira S, Ravelli A, Pistorio A, et al. Proxy-reported health-related quality of life of patients with juvenile idiopathic arthritis: the Pediatric Rheumatology International Trials Organization multinational quality of life cohort study. Arthritis Rheum 2007;57(1):35-43.
61. Selvaag AM, Flato B, Lien G, et al. Measuring health status in early juvenile idiopathic arthritis: determinants and responsiveness of the child health questionnaire. J Rheumatol 2003;30(7):1602-10.
62. Sawyer MG, Whitham JN, Roberton DM, et al. The relationship between health-related quality of life, pain and coping strategies in juvenile idiopathic arthritis. Rheumatology (Oxford) 2004;43(3):325-330.
63. Bazso A, Consolaro A, Ruperto N, et al. Development and testing of reduced joint counts in juvenile idiopathic arthritis. J Rheumatol 2009;36(1):183-90.
64. Bekkering WP, ten Cate R, van Suijlekom-Smit LW, et al. The relationship between impairments in joint function and disabilities in independent function in children with systemic juvenile idiopathic arthritis. J Rheumatol 2001;28(5):1099-105.
65. Magni-Manzoni S, Cugno C, Pistorio A, et al. Responsiveness of clinical measures to flare of disease activity in juvenile idiopathic arthritis. Clin Exp Rheumatol 2005;23(3):421-5.
66. Ruperto N, Ravelli A, Falcini F, et al. Responsiveness of outcome measures in juvenile chronic arthritis. Italian Pediatric Rheumatology Study Group. Rheumatology (Oxford) 1999;38(2):176-80.
67. Ruperto N, Ravelli A, Migliavacca D, et al. Responsiveness of clinical measures in children with oligoarticular juvenile chronic arthritis. J Rheumatol 1999;26(8):1827-30.
68. Moretti C, Viola S, Pistorio A, et al. Relative responsiveness of condition specific and generic health status measures in juvenile idiopathic arthritis. Ann Rheum Dis 2005;64(2):257-61.
69. Filocamo G, Sztajnbok F, Cespedes-Cruz A, et al. Development and validation of a new short and simple measure of physical function for juvenile idiopathic arthritis. Arthritis Rheum 2007;57(6):913-20.
70. Brunner HI, Klein-Gitelman MS, Miller MJ, et al. Health of children with chronic arthritis: relationship of different measures and the quality of parent proxy reporting. Arthritis Rheum 2004;51(5):763-73.
71. Consolaro A, Vitale R, Pistorio A, et al. Physicians' and parents' ratings of inactive disease are frequently discordant in juvenile idiopathic arthritis. J Rheumatol 2007;34(8):1773-6.
72. Sztajnbok F, Coronel-Martinez DL, Diaz-Maldonado A, et al. Discordance between physician's and parent's global assessments in juvenile idiopathic arthritis. Rheumatology (Oxford) 2007;46(1):141-5.
73. Lurati A, Pontikaki I, Teruzzi B, et al. A comparison of response criteria to evaluate therapeutic response in patients with juvenile idiopathic arthritis treated with methotrexate and/or anti-tumor necrosis factor alpha agents. Arthritis Rheum 2006;54(5):1602-7.
74. Giannini EH, Ruperto N, Ravelli A, et al. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40(7):1202-9.
75. Ruperto N, Ravelli A, Falcini F, et al. Performance of the preliminary definition of improvement in juvenile chronic arthritis patients

treated with methotrexate. Italian Pediatric Rheumatology Study Group. Ann Rheum Dis 1998;57(1):38-41.
76. Sawyer MG, Carbone JA, Whitham JN, et al. The relationship between health-related quality of life, pain, and coping strategies in juvenile arthritis--a one year prospective study. Qual Life Res 2005;14(6):1585-98.
77. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28(7):632-42.
78. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 1991;12(4 Suppl):142S-158S.
79. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40(2):171-8.
80. Cohen J. Statistical power for the behavioral sciences (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
81. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986;39(11):897-906.
82. Magni-Manzoni S, Ruperto N, Pistorio A, et al. Development and validation of a preliminary definition of minimal disease activity in patients with juvenile idiopathic arthritis. Arthritis Rheum 2008;59(8):1120-7.
83. Magni-Manzoni S, Pistorio A, Labo E, et al. A longitudinal analysis of physical functional disability over the course of juvenile idiopathic arthritis. Ann Rheum Dis 2008;67(8):1159-1164.

Abbreviations

ACR	American College of Rheumatology
ACR Pediatric	American College of Rheumatology Pediatric Response Criteria
AHRQ	Agency for Healthcare Research and Quality
AJC	Active joint count
CD	Cluster of differentiation
CHAQ	Childhood Health Assessment Questionnaire
CHAQ-DI	Childhood Health Assessment Questionnaire Disability Index
CHQ	Child Health Questionnaire
CHQ PhS	Child Health Questionnaire physical score
CHQ PsS	Child Health Questionnaire psychosocial score
CI	Confidence interval
DMARD(s)	Disease-modifying anti-rheumatic drug(s)
EPC	Evidence-based Practice Center
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
Fc	Fragment crystallizable
FDA	U.S. Food and Drug Administration
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HAQ	Stanford Health Assessment Questionnaire
IgM	Immunoglobulin M
IL	Interleukin
ILAR	International League of Associations for Rheumatology
IVIG	Intravenous immunoglobulin
JCA	Juvenile chronic arthritis
JIA	Juvenile idiopathic arthritis
JRA	Juvenile rheumatoid arthritis
LROM	Limited range of motion
MDA	Minimal disease activity
MeSH	Medical Subject Headings
NSAID(s)	Non-steroidal anti-inflammatory drug(s)
PedsQL	Pediatric Quality of Life Inventory
PedsQL-RM	Pediatric Quality of Life Inventory Rheumatology Module
PGA	Physician global assessment of disease activity
PGW	Parent/patient global assessment of well-being
PICOTS	Population, interventions, comparators of interest, outcomes, timing, and settings
RCT	Randomized controlled trial
ROC	Receiver operating characteristic
RR	Risk ratio
SAARD(s)	Slow-acting anti-rheumatic drug(s)
SRC	Scientific Resource Center
T-cell/-lymphocyte	Thymus cell/lymphocyte
TEP	Technical expert panel
TNF	Tumor necrosis factor