Decisional Dilemmas in Discontinuing Prolonged Disease-Modifying Treatment for Multiple Sclerosis
Decisional Dilemmas in Discontinuing Prolonged Disease-Modifying Treatment for Multiple Sclerosis

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

Contract No. 290-2012-00016-I

Prepared by:
Minnesota Evidence-based Practice Center
Minneapolis, MN

Investigators:
Mary Butler, Ph.D., M.B.A.
Mary L. Forte, Ph.D., D.C.
Natalie Schwehr, M.Ac., L.Ac.
Adam Carpenter, M.D.
Robert L. Kane, M.D.

AHRQ Publication No. 15-EHC012-EF
April 2015
This report is based on research conducted by the Minnesota Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2012-00016-I). The findings and conclusions in this document are those of the authors, 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.

AHRQ or U.S. Department of Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies, may not be stated or implied.

This report may periodically be assessed for the currency of conclusions. If an assessment is done, the resulting surveillance report describing the methodology and findings will be found on the Effective Health Care Program Web site at www.effectivehealthcare.ahrq.gov. Search on the title of the report.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

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

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies.

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 can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see www.effectivehealthcare.ahrq.gov/reference/purpose.cfm.

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. 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 email list to learn about new program products and opportunities for input.

We welcome comments on this systematic review. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Robert G. Kronick, Ph.D. 
Director 
Agency for Healthcare Research and Quality

David Meyers, M.D. 
Acting Director 
Center for Evidence and Practice Improvement 
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H. 
Director, EPC Program 
Center for Evidence and Practice Improvement 
Agency for Healthcare Research and Quality

Suchitra Iyer, Ph.D. 
Task Order Officer 
Center for Evidence and Practice Improvement 
Agency for Healthcare Research and Quality
Acknowledgments

The authors gratefully acknowledge Suchitra Iyer, Melissa McPherson, Marilyn Eells, and Jeannine Ouellette for their contributions to this project; this report is better for them.

Key Informants

In designing the study questions, the EPC consulted several Key Informants who represent the end-users of research. The EPC sought the Key Informant input on the priority areas for research and synthesis. Key Informants are not involved in the analysis of the evidence or the writing of the report. Therefore, in the end, study questions, design, methodological approaches, and/or conclusions do not necessarily represent the views of individual Key Informants.

Key Informants must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any conflicts of interest.

The list of Key Informants who participated in developing this report follows:

Thomas Getchius, Ph.D.
Associate Director, Clinical Practice
American Academy of Neurology
Minneapolis, MN

Gary S. Gronseth, M.D.
Department of Neurology
University of Kansas
Kansas City, KS

Michael Kaufman, M.D.
Knoxville Neurology Clinic
University of Tennessee (Knoxville) MS Center
Knoxville, TN

Nicholas LaRocca, Ph.D.
National Multiple Sclerosis Society
New York, NY

Brant Oliver, Ph.D., M.S., M.P.H., A.P.R.N.-B.C.
MGH Institute of Health Professions
Boston MA

Alex Duart Rae-Grant, M.D.
Mellen Center for Multiple Sclerosis
Cleveland Clinic Main Campus
Cleveland, OH

Lisa Taylor Skutnik, P.T., M.A.
National Multiple Sclerosis Society
New York, NY

Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.
The list of Technical Experts who participated in developing this report follows:

*Gary S. Gronseth, M.D.
Department of Neurology
University of Kansas
Kansas City, KS

*Michael Kaufman, M.D.
Knoxville Neurology Clinic
University of Tennessee (Knoxville) MS Center
Knoxville, TN

*Nicholas LaRocca, Ph.D.
National Multiple Sclerosis Society
New York, NY

*Brant Oliver, Ph.D., M.S., M.P.H., A.P.R.N.-B.C.
Assistant Professor, School of Nursing, MGH Institute of Health Professions
Boston, MA
Faculty Senior Scholar
Department of Veterans Affairs National Quality Scholars Fellowship Program
White River Junction, VT

Alex Duart Rae-Grant, M.D.
Mellen Center for Multiple Sclerosis
Cleveland Clinic Main Campus
Cleveland, OH

*Also served as Peer Reviewer.

**Peer Reviewers**

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report do not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential nonfinancial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

The list of Peer Reviewers follows:

Christopher Luzzio, M.D.
Department of Neurology
University of Wisconsin
Madison, WI

Sarah Morrow, M.D., M.S., FRCPC
Clinical Neurological Sciences
Western University
London, Ontario, Canada

Anthony Reder, M.D.
University of Chicago Medicine
Chicago, IL

Andrew J. Solomon, M.D.
Department of Neurology Sciences
University of Vermont
Burlington, VT
Decisional Dilemmas in Discontinuing Prolonged Disease-Modifying Treatment for Multiple Sclerosis

Structured Abstract

Objective. We conducted a systematic review to examine the long-term consequences of discontinuing disease-modifying treatment (DMT) for multiple sclerosis (MS) by examining the long-term benefits and harms, and the reasons for discontinuing treatment. We also examined the evidence for people’s values, beliefs, and preferences regarding discontinuing DMT.

Data sources. We searched Medline®, PsycInfo®, Scopus, and the Cochrane Clinical Trials Registry through August 2014 plus reference lists of included studies and recent systematic reviews.

Methods. Two investigators screened abstracts and full texts of identified references for eligibility. Eligible studies included studies of over 3 years that examined Food and Drug Administration–approved DMTs compared with placebo, other active DMT, or no DMT for adults with clinically isolated syndrome or MS in outpatient settings for patient-centered outcomes. We excluded studies of mitoxantrone, since it has a maximum lifetime dosage. Timing was relaxed for women who were considering pregnancy or already pregnant or patients discontinuing natalizumab due to risk factor changes. We extracted data, assessed risk of bias of individual studies, and evaluated strength of the body of evidence for each comparison and outcome. We also evaluated, using Technical Brief methods, studies of any design that examined individuals’ attitudes, values, and preferences for discontinuing treatments and health states, or factors and processes patients with MS and clinicians use in shared decisionmaking.

Results. We identified 27 unique studies with discontinuation information: 16 of these contained complete information to allow full analysis of long-term benefits and harms. Evidence was insufficient for long-term benefits of DMTs for secondary progressive MS patients and most outcomes for relapsing-remitting MS (RRMS) patients. Low-strength evidence suggests higher long-term all-cause survival for treatment-naïve RRMS patients who did not delay starting interferon beta-1b by 2 years and used DMTs for a longer duration than for those who started later. Low-strength evidence suggests that interferon did not change RRMS patients’ disability progression. Limited low-strength evidence suggests that long-term harms do not differ from short-term harms. The majority of discontinuation tends to occur within 2 to 3 years. Another 25 unique studies provided intrapersonal, interpersonal, and shared decisionmaking information. No study directly asked why people may be reluctant to discontinue when treatment no longer seems effective; taken as a whole, the literature set provides some insight. The preferences literature underscores the complexity of the topic and the processes underlying decisionmaking.

Conclusions. MS patients and providers have little information to guide decisions to discontinue DMT.
## Contents

**Executive Summary** ............................................................ES-1

**Introduction** ..............................................................................................................1
- Background...............................................................................................................1
- Condition..................................................................................................................1
- Treatment Strategies – and Their Discontinuation ..................................................2
- Scope and Key Questions .........................................................................................3
- Scope of the Review ..................................................................................................3
- Key Questions............................................................................................................4
- Organization of This Report .....................................................................................6

**Methods** ....................................................................................................................7
- Topic Refinement and Review Protocol .....................................................................7
- Role of the AHRQ Task Order Officer .......................................................................7
- Analytic Framework ...................................................................................................7
- Literature Search Strategy ...........................................................................................8
- Search Strategy ...........................................................................................................8
- Inclusion and Exclusion Criteria ................................................................................9
- Study Selection .........................................................................................................10
- Data Extraction .........................................................................................................10
- Risk of Bias Assessment of Individual Studies .........................................................10
- Data Synthesis ..........................................................................................................11
- Strength of the Body of Evidence ..............................................................................11
- Applicability ..............................................................................................................12

**Results** .....................................................................................................................13
- Results of Literature Searches ....................................................................................13
- KQ1. Effectiveness of Discontinuing Disease-Modifying Treatments .......................14
  - Description of Included Studies ...........................................................................14
  - KQ1a. Evidence for Benefits for Continuing Treatment Versus Discontinuing ......17
  - KQ1b. Evidence for Long-Term Harms ..................................................................22
  - Natalizumab Discontinuation or Drug Holiday .....................................................24
  - Pregnancy Drug Holiday .......................................................................................28
  - KQ1c. Reasons for Discontinuation of Disease-Modifying Treatments Reported in Long-Term Observational Cohort Studies ........................................30
- KQ2. Individuals’ Values, Beliefs, and Preferences Regarding Discontinuing Disease-Modifying Treatments .................................................................31
  - Description of Included Studies ...........................................................................31
  - Overarching Key Point ............................................................................................31
  - KQ2a. Patients’ and Providers’ Preferences for Discontinuation of Disease-Modifying Treatments .................................................................31
  - KQ2b. Patients’ and Providers’ Preferences for Participation in Shared Decisionmaking To Discontinue Disease-Modifying Treatments ..............................43
Appendix D. Risk of Bias for Included KQ1 Studies
Executive Summary

Background

Multiple sclerosis (MS) is a variably debilitating disease characterized by demyelination (deterioration of the protective myelin sheaths covering nerve cell processes in the brain and spinal cord) and axon loss within the central nervous system. The lesions created by the myelin destruction and resulting scar tissue interfere with normal transmission along nerve fibers within the brain and to and from the brain. This results in classic symptoms associated with MS. The condition affects 2.5 million individuals worldwide and approximately 400,000 in the United States. About 40 percent of people with MS receive some form of disability income. Twice as many women as men are affected, and diagnosis usually occurs between the ages of 20 and 50. Symptoms and disease course are highly individual, depending on where the lesions occur within the central nervous system and the type of MS. Clinically definite MS types include the following:

- **Relapsing-remitting MS (RRMS)** is the most common form, affecting approximately 85 percent of patients. Patients typically are diagnosed in their 20s or 30s. Neurologic symptoms of a relapse typically develop over a course of days, stabilize, and spontaneously improve. However, over time permanent disability often accrues, with further relapses. Many patients with RRMS eventually transition to secondary progressive MS (below). Estimates of the median time from RRMS onset to this transition range from 15 to 29 years.

- **Secondary progressive MS (SPMS)** is characterized by worsening disability with or without relapses. Patients may have exacerbations, but the trend over time is a relatively steady progression of disease and disability.

- **Primary progressive MS (PPMS)** represents about 15 percent of patients and affects women and men about equally. This form has the worst prognosis and is characterized by gradual and progressive worsening of function without distinct relapses.

- **Progressive relapsing MS (PRMS)** affects about 5 percent of patients. This form is usually diagnosed first as PPMS due to a steady worsening of functioning and changed to PRMS when the patient experiences a relapse. Recently, a recommendation has been made to eliminate PRMS as a type, classifying these patients as having PPMS.

People with clinically isolated syndrome (CIS), a first neurologic episode consistent with an MS relapse, may or may not go on to develop MS. CIS involves neurologic symptoms such as vision loss, numbness, or weakness that last at least 24 hours and are caused by inflammation or demyelination in one (monofocal) or more (multifocal) sites in the central nervous system. In a cohort of 107 CIS patients followed for 20 years, 60 patients with three or more lesions (seen via magnetic resonance imaging [MRI]) converted to definite MS, while only 7 with normal baseline MRI converted.

MS cannot be cured with current therapies. Disease-modifying treatments (DMTs) comprise immunomodulating and immunosuppressant medications aimed at slowing the progression of MS and improving quality of life. The working hypothesis is that reducing or preventing new lesions and their sequelae slows the worsening of the disease. DMTs currently approved by the Food and Drug Administration (FDA) for RRMS include interferon (IFN) beta-1a and -1b (some formulations also approved for CIS), glatiramer acetate, mitoxantrone (also approved for SPMS and PRMS), natalizumab, fingolimod, and dimethyl fumarate.
A 2013 Cochrane overview review and network analysis of 44 2- to 3-year trials of DMTs for MS found moderate- to high-quality evidence that DMTs are effective against recurrence of relapses in RRMS during the first 24 months of treatment compared with placebo. The network analysis ranked natalizumab as the most effective drug, followed in order by IFNbeta-1a (Rebif®), mitoxantrone, glatiramer acetate (Copaxone®), and IFNbeta-1b (Betaseron®). Confidence in the evidence dropped to moderate for direct comparisons of mitoxantrone or IFNbeta-1b versus placebo and very low for glatiramer acetate versus placebo. Further, natalizumab and IFNbeta-1b were more effective than IFNbeta-1a in reducing the number of RRMS participants with disease progression, as measured with surrogate markers. In patients with progressive MS, both pairwise and network analysis found that no DMT analyzed prevented disability progression over 2 or 3 years. The overview and network analysis were too recent to include the newest approved drugs, such as fingolimod or dimethyl fumerate.

Unfortunately, the efficacy of MS treatments appears to correlate with the frequency and severity of side effects. The injectable treatments, the IFN drugs and glatiramer acetate, were modestly efficacious and side effects were tolerable by many patients. Mitoxantrone, an escalation medication, has a lifetime maximum dosage due to cardiotoxicity and risks of leukemia. Natalizumab, the first monoclonal antibody approved for treating MS, can induce the potentially fatal brain infection progressive multifocal leukoencephalopathy (PML). Risk for PML increases with natalizumab use longer than 2 years, anti-JC virus antibody status, and prior use of immunosuppressive agents. People taking natalizumab may take a drug holiday or discontinue use completely if their risk for PML increases, assessed by a positive test for the anti-JC virus antibody status.

Women considering pregnancy face special considerations for drug holidays. There are no class A drugs (drugs safe for use during pregnancy according to the FDA) for MS. Women and their physicians must weigh the possible risks of DMT exposure to the unborn fetus against the maternal risk of disease progression if they discontinue DMT.

The optimal duration of DMT use remains an open and controversial question. Many patients do not use these medications throughout their entire life after diagnosis. However, with few exceptions (such as natalizumab use or intended pregnancy), patients who opt for DMT for MS may end up using it for several years to decades, as long as they tolerate the treatment and the DMT seems effective. Patients may switch between DMTs in order to find one that is more effective or more tolerable, and studies have found high rates of switching between drugs. Some patients cannot tolerate any of the DMTs, but if a tolerable drug regime is determined, treatment generally continues until the individual reaches a disease stage where DMTs are no longer considered to be helping. Such a point may be reached when it is determined that a person is nonresponsive to the medication due to disease progression. Determining when DMT is no longer helpful is challenging. Thus, major questions of interest are whether or not DMTs for MS alter the natural history of the disease in the long run and when (if ever) to discontinue DMT. The related question addresses the influence of patient values, beliefs, and preferences regarding discontinuing DMTs. Such information should support clinicians, patients, consumer advocates, and other decisionmakers on decisions to discontinue treatment.

**Scope and Key Questions**

This review examines the long-term (more than 3 years) consequences of continuing or discontinuing DMT. We looked for evidence that directly assessed discontinuing versus continuing DMT and also evidence for long-term (more than 3 years) benefits and harms for
either continuing or discontinuing, since that decision can be informed by the benefits or harms
directly linked to either course of action. This information would extend understanding beyond
the short-term trials examined in the 2013 Cochrane review. We were also interested in the
reasons for discontinuing treatment reported in the long-term studies.

We concentrated on outcomes that patients notice or factor directly into their
decisionmaking, such as relapse rates and changes in disability level, rather than intermediate
outcomes such as lab tests for neutralizing antibodies. MRI to identify MS-related lesions has
been shown to correlate with short-term relapse rates (6 months to 2 years).7 However, long-term
MRI followup results as a surrogate marker for relapse rates or, more importantly, disease
progression, currently lack evidence.10,11 Further, short-term MRI followup as a predictor of
long-term disability progression answers a different research question—i.e., does short-term
treatment affect long-term outcomes—than the research questions asked for this review. Thus,
we did not use MRI results as a long-term outcome in this review. However, we included MRI
results as a short-term outcome in the subset of patients discontinuing natalizumab due to risk of
PML.

People with MS commonly switch between the available DMTs, depending on tolerance,
presence of adverse effects, and perceived helpfulness of the treatment. The pertinent clinical
question for switching medications is how to define the threshold of disease activity for changing
medications. This important question is qualitatively different from that of when, if ever, to stop
DMT completely. To adequately address the question of when to switch medications will likely
require a review of both short- and long-term research. Therefore, questions related to switching
between DMTs are outside the scope of this review.

We synthesized the evidence in the published literature to address the following two Key
Questions (KQs):

KQ1: What are the consequences of discontinuing disease-modifying
treatments in adult patients?
   a. What is the evidence for benefits for continuing versus discontinuing treatment?
   b. What is the evidence for long-term harms?
   c. What reasons for discontinuation of disease-modifying treatments have been reported in
      long-term observational cohort studies?

KQ2: What are individual values, beliefs, and preferences regarding
discontinuing disease-modifying treatments?
   a. What are patient and provider preferences for discontinuation of disease-modifying
treatments?
   b. What are patient and provider preferences for participation in shared decisionmaking to
      discontinue disease-modifying treatments?

Figure A provides a conceptual framework that links the KQs. At the top it depicts the logic
path both physicians and patients must travel when considering DMT:

- Does it work?
- What drug should I start with?
- When should I switch to a new drug and what should that drug be?
- When should I discontinue DMT?
This logic path describes the context within which patients and clinicians consider clinical factors—tolerability of the medication, disease characteristics at the time of discontinuation (relapses, progression, MRI activity), risk of ongoing disease treatment, other impediments to continued medication use (difficulty in obtaining, injecting, or ingesting, cost, etc.)—and make decisions about DMT or, in the case of this review, discontinuation (KQ1). The lower part of the figure, the conceptual basis for KQ2, depicts the progression from an individual's internal decision context and process (such as preferences, values, knowledge, beliefs, and cognitive behaviors and habits) to an interpersonal decision context and processes between the physician and patient. The overlapping ovals representing the clinician and the patient indicate information shared between the two parties versus information and other cognitive processes specific to one individual. Any overlap depends in part on the level of sophistication a patient brings to the decisionmaking process and in part on how well a physician understands a patient’s beliefs, values, goals, and preferences. For example, a patient newly diagnosed with MS in the novice phase of learning about MS would likely have a smaller overlap. The interaction between the physician and patient results in decisions that can vary in their level of concordance.

Figure A. Conceptual framework for Key Questions

KQ = Key Question
Figure B provides an analytic framework describing the treatment path and long-term benefits and harms of continuing versus discontinuing DMT for KQ1.

**Figure B. Analytic framework for discontinuing disease-modifying treatments for MS**

- **KQ2**: Individual values, beliefs, and preferences for discontinuing DMT
- **KQ1a**: Discontinuing vs. continuing long-term DMT
- **Final health outcomes**
  - Relapse rate
  - Change in disability
  - Progression of disease
  - Time to sustained disease progression
- **KQ1b**: Adverse effects of continuing or discontinuing long-term DMT
- **KQ1c**: Other reasons provided for discontinuing treatment

DMT = disease-modifying treatment; KQ = Key Question; MS = multiple sclerosis

**Methods**

The methods for this review follow the methods suggested in the Agency for Healthcare Research and Quality “Methods Guide for Effectiveness and Comparative Effectiveness Reviews” (available at www.effectivehealthcare.ahrq.gov). We used Technical Brief methods for KQ2. A complete description of the methods can be found in the full report. All methods and analyses were determined a priori.

**Literature Search Strategy**

We used bibliographic databases to identify publications on randomized controlled trials (RCTs), systematic reviews, and observational studies with control groups published from 1990 to August 2014 that enrolled adults with CIS or MS. Relevant bibliographic databases for this topic include MEDLINE®, the Cochrane Central Register of Controlled Trials (CENTRAL), PsycInfo®, and Scopus. We supplemented bibliographic database searches with backward citation searches of highly relevant systematic reviews.

**Eligibility**

For KQ1, we included studies from 1990 through August 2014 that examined patient-centered outcomes for patients with CIS or MS in outpatient settings using FDA-approved (through August 2014) DMT compared with placebo, other active DMT, or no DMT. We excluded studies of pediatric MS patients, studies of mitoxantrone (since it has a maximum lifetime dosage), and studies with 3-year or less followup. However timing was relaxed for...
women who were considering pregnancy or were pregnant, or patients discontinuing natalizumab due to changes in risk of PML.

For KQ2, we included studies of any design that examined individuals’ attitudes, values, preferences for discontinuing treatments and health states, perceptions of risk and seriousness of health states, or factors and processes patients with MS and clinicians use in shared decisionmaking. Two independent investigators independently determined study eligibility and resolved disagreements through discussions (possibly with a third adjudicator) until consensus was achieved. Study selection involved an extensive full-text review process to identify adult subgroups, since subgroup reporting was commonly not evident in titles and abstracts.

Data Extraction

We extracted data from included studies into standardized evidence tables. Extracted data included relevant population, intervention, baseline, and outcomes data on the adult subgroups of interest. Initial data abstraction was quality checked by a second investigator.

Risk-of-Bias Assessment of Individual Studies

The risk of bias of eligible studies was assessed by two independent investigators using instruments specific to each study design for KQ1. The two investigators consulted to reconcile any discrepancies in overall risk-of-bias assessments and, when needed, a third investigator was consulted to reconcile the summary judgment. For KQ1, we developed an instrument to assess risk of bias for observational studies using the RTI Observational Studies Risk of Bias and Precision Item Bank.13 We selected items most relevant in assessing risk of bias for this topic, including participant selection, ascertainment, attrition, performance, and appropriateness of analytic methods. Following Technical Brief methods, risk of bias was not assessed for KQ2.

Data Synthesis

For KQ1, we summarized the results into evidence tables and qualitatively synthesized evidence for comparisons for specific disease-modifying medications, unique populations, duration of DMT, length of study followup, and outcomes. We used the best of the evidence provided by the identified observational literature.14 So, while all identified articles underwent abstraction, only the best evidence, based on those studies closest to an “ideal” study design15 (those studies with the lowest risk of bias), are included in the evidence synthesis.

For KQ2, we summarized the results into evidence tables and conducted a qualitative synthesis. We grouped the literature by mapping the included studies to the conceptual framework (Figure A) and analyzed the study findings for emergent patterns in patient perspectives, clinician perspectives, and clinician/patient interpersonal interactions.

Strength of the Body of Evidence

The overall strength of evidence for selected outcomes for KQ1 (relapse rate, change in disability, progression of disease, time to sustained disease progression) within each comparison was evaluated based on four required domains: (1) study limitations (internal validity); (2) directness (single direct link between intervention and outcome); (3) consistency (similarity of effect direction and size); and (4) precision (degree of certainty around an estimate).16 A fifth domain, reporting bias, was assessed when strength of evidence based on the first four domains

ES-6
was moderate or high. Based on study design and conduct, risk of bias was rated as low, medium, or high. Consistency was rated as consistent, inconsistent, or unknown/not applicable (e.g., single study). Directness was rated as either direct or indirect. Precision was rated as precise or imprecise. Other factors that may be considered in assessing strength of evidence include dose-response relationship, the presence of confounders, and strength of association. Based on these factors, the overall evidence for each outcome was rated as follows:

- **High**—Very confident that estimate of effect lies close to true effect. Few or no deficiencies in body of evidence; findings believed to be stable.
- **Moderate**—Moderately confident that estimate of effect lies close to true effect. Some deficiencies in body of evidence; findings likely to be stable, but some doubt.
- **Low**—Limited confidence that estimate of effect lies close to true effect; major or numerous deficiencies in body of evidence. Additional evidence necessary before concluding that findings are stable or that estimate of effect is close to true effect.
- **Insufficient**—No evidence, unable to estimate an effect, or no confidence in estimate of effect. No evidence is available or the body of evidence precludes judgment.

Following Technical Brief methods, strength of evidence was not assessed for KQ2. This KQ was approached in a hypothesis-generating manner.

**Applicability**

Applicability of studies was determined according to the PICOTS (population, intervention, comparison, outcomes, timing, setting) framework. Study characteristics that may affect applicability include, but are not limited to, type of MS or CIS, unobserved differences in patient preferences, or country in which treatment is provided, given differences in international regulations and treatment preferences.

**Results**

We identified 4,359 unique citations in searching from 1990 to August 2014. After excluding articles at the title and abstract phase, full texts of 198 articles were reviewed to determine final inclusion. Seven articles were added through hand-search. Of the 61 articles retained for KQ1, 11 were specific to discontinuing natalizumab due to increased risk and 12 were specific to discontinuing due to pregnancy. Of the remaining 38 articles comprising 27 unique studies, only 16 studies contained complete information to allow for full analysis. All 38 articles were reviewed for information on reported reasons for discontinuation. For KQ2, 30 articles comprising 27 unique studies were included. Detailed tables and synthesis can be found in the full report.

**KQ1a. Benefits of Continuing Versus Discontinuing DMT**

The key points for KQ1a are as follows:

- No studies directly assessed the consequences of continuing versus discontinuing DMT in comparable populations.
- Low-strength evidence from one study with moderate risk of bias suggests that, for RRMS patients, long-term all-cause survival is higher for treatment-naïve patients who did not delay starting IFNb-1b by 2 years and used DMT for a longer duration than those who started later.
Low-strength evidence from one study with moderate risk of bias suggests that IFN use did not change disability progression for RRMS patients.

Insufficient evidence was available to assess long-term benefits of DMT for SPMS patients and most outcomes for RRMS patients. Except for those noted above, studies were high risk of bias and had small sample sizes, and reported effects were small in magnitude.

Results are summarized in Table A.

<table>
<thead>
<tr>
<th>DMT and Author</th>
<th>Type of MS at Baseline</th>
<th>Median or Mean Years to Final Assessment</th>
<th>All-Cause Mortality</th>
<th>Convert to SPMS</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNbeta-1b: Goodin, 2012</td>
<td>RRMS</td>
<td>21</td>
<td>All-cause mortality: 250 mg arm vs. placebo—HR, 0.532 (95% CI, 0.31 to 0.90); 50 mg arm vs. placebo—HR, 0.54 (95% CI, 0.32 to 0.92) Favors treatment</td>
<td>NR</td>
<td>Low</td>
</tr>
<tr>
<td>IFNbeta, mixed: Shirani, 2012</td>
<td>RRMS</td>
<td>4.5–10.5</td>
<td>NR</td>
<td>Time to sustained EDSS 6: no difference from contemporary control (HR, 1.30; 95% CI, 0.92 to 1.83) or historical control (HR, 0.77; 95% CI, 0.58 to 1.02) Favors treatment</td>
<td>NR</td>
</tr>
</tbody>
</table>

CI = confidence interval; DMT = disease-modifying treatment; EDSS = Expanded Disability Status Scale; HR = hazard ratio; IFN = interferon; NR = not reported; MS = multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis

No studies directly assessed continuing versus discontinuing DMT in comparable populations. We therefore turned to literature examining benefits for continuing DMT long term. Variation among the included studies on long-term benefits of DMT for patient populations, interventions, outcome measurements, and timeframes precluded meaningful pooling. Only two studies provide a low strength of evidence for two benefit outcomes measured long term for IFNs.

One study with moderate risk of bias examined all-cause mortality over a 21-year period for 366 patients who had enrolled in an RCT (98.4% of the original RCT participants) testing IFNbeta-1b for treatment-naïve RRMS patients in 11 clinics in North America. The study’s strength lies in the nearly complete followup of patients and the objective outcome measure. Participants in the two treatment arms (50 mg and 250 mg) showed lower all-cause mortality compared with the placebo arm. The survival rate for the placebo arm was consistent with survival rates reported in MS natural history studies. Median treatment duration for the three groups ranged from 7 years for the placebo group to 14 years for the 50 mg arm and 12 years for the 250 mg arm. Patients assigned to placebo had both later starts and shorter exposure to DMTs. Thus, the study cannot distinguish between the effects of early use and the effects of long-term use.
One study with moderate risk of bias examined the association between IFNbeta use and progression to a sustained Expanded Disability Status Scale (EDSS) score of 6 for 2,656 RRMS patients in Canada. Three arms were used: a treatment cohort followed for 5.1 years, a contemporary cohort followed 4 years, and a historical cohort (drawn from the pre-IFN period) followed 10.8 years. The strength of this study lies in the almost complete capture of MS patients, since patients were unable to obtain DMTs other than from the participating clinic, and the multiple statistical approaches used to test for association, including use of comorbidities (Charlson score) and socioeconomic status along with age, sex, disease duration, and EDSS. Propensity score adjustments did not substantially change the results. The study did not find statistically significant differences in hazard ratios for reaching a sustained EDSS score of 6 for either contemporary or historical cohort comparisons.

Insufficient evidence exists to address long-term benefits for glatiramer acetate, teriflunomide, and natalizumab for either RRMS or SPMS, as well as important MS outcomes for IFNbeta for RRMS beyond all-cause mortality or 5-year disability progression.

**KQ1b. Evidence for Harms**

The key points for KQ1b are as follows:

- Limited low-strength evidence suggests that harms for injectable DMTs do not differ between short term (2-3 years) and long term (up to 16 years for IFN, 22 years for glatiramer acetate, and 8.5 years for teriflunomide).
- The majority of discontinuation tends to occur in the short term (2-3 years from start).
- Broad variation in harms reporting precludes informative aggregation and summary.
- Because of high risk of bias and small sample sizes, evidence is insufficient for whether rebound after discontinuing natalizumab exists.
- Because of high risk of bias and small sample sizes, evidence is insufficient to address the risks of fetal exposure to DMT during pregnancy in women with MS or the risks to the mother from the drug holiday.

Results are summarized in Table B.

<table>
<thead>
<tr>
<th>DMT</th>
<th>Number of Studies; Total N; Followup</th>
<th>Any Adverse Event</th>
<th>At Least 1 Serious Adverse Event</th>
<th>Treatment Discontinuation for Adverse Event</th>
<th>Comparator Groups</th>
<th>Reported Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNbeta-1a</td>
<td>1 N = 429 4</td>
<td>Most common AEs: injection site reactions, headache, flulike symptoms</td>
<td>NR</td>
<td>NR</td>
<td>No</td>
<td>Long-term events do not differ from short-term events.</td>
</tr>
<tr>
<td>IFNbeta-1b</td>
<td>2 N = 746 5–16</td>
<td>Most common AEs: injection site reactions, depression, flulike symptoms, headache</td>
<td>21% to 24%</td>
<td>Discontinuation rates “high” but numbers not reported</td>
<td>No</td>
<td>Long-term events do not differ from short-term events. Frequency declined over 16 years in continuers.</td>
</tr>
</tbody>
</table>
### Table B. Harms reported from unique studies included in the analytic set (continued)

<table>
<thead>
<tr>
<th>DMT</th>
<th>Number of Studies; Total N; Followup</th>
<th>Any Adverse Event</th>
<th>At Least 1 Serious Adverse Event</th>
<th>Treatment Discontinuation for Adverse Event</th>
<th>Comparator Groups</th>
<th>Reported Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNbeta, mixed&lt;sup&gt;9,23,24&lt;/sup&gt;</td>
<td>3 N = 587 4–8</td>
<td>Most common AEs: injection site reactions, depression, flulike symptoms, headache</td>
<td>NR</td>
<td>3% during long-term followup; discontinuation for serious AE more likely to happen early in treatment course (1 year)</td>
<td>No</td>
<td>Headache more likely for IFNbeta-1a; injection site reactions more likely for IFNbeta-1b. No other differences between type of IFNbeta. Majority of discontinuations occur early/short term.</td>
</tr>
<tr>
<td>IFNbeta, mixed&lt;sup&gt;25&lt;/sup&gt;, SPMS</td>
<td>1 N = 146 5</td>
<td>NR</td>
<td>NR</td>
<td>3.4%, although timing is not clear</td>
<td>No</td>
<td>Majority of discontinuations occur early/short term.</td>
</tr>
<tr>
<td>Glatiramer acetate&lt;sup&gt;26-28&lt;/sup&gt;</td>
<td>3 N = 483 4–22</td>
<td>Only 1 reported overall rate: 87.3% Most common AE: injection site reactions</td>
<td>NR</td>
<td>Only 1 reported overall rate: 4.9% in long-term extension</td>
<td>No</td>
<td>Majority of discontinuations occur early/short term. Long-term events do not differ from short-term events.</td>
</tr>
<tr>
<td>Teriflunomide&lt;sup&gt;29&lt;/sup&gt;</td>
<td>1 N = 147 8.5</td>
<td>98% of 7 mg dose and 100% of 14 mg dose experienced treatment-emergent AE</td>
<td>36% of 7 mg dose and 29% of 14 mg dose</td>
<td>13.6% of 7 mg dose and 13.6% of 14 mg dose</td>
<td>1 comparison to general population rates for cancer</td>
<td>Long-term events do not differ from short-term events.</td>
</tr>
</tbody>
</table>

AE = adverse event; IFN = interferon; NR = not reported; SPMS = secondary progressive multiple sclerosis

Eleven of the 16 unique studies reported harms in enough detail for abstraction. Only one of the studies was moderate risk of bias; all others were rated as high risk of bias.

The included studies used a wide range of reporting methods and adverse event categories that precluded simple aggregation over the studies. The most commonly reported adverse events were injection site reactions, flulike symptoms, depression, and headache. Serious adverse events were generally not reported, although two studies gave rates of about 25 percent to 30 percent of participants. Discontinuations tended to occur during the first or second year of the study. When reported, discontinuation rates during long-term followup were low, about 3 to 4 percent, but rates due to adverse events were not separate from total discontinuation rates, which would also include perceived lack of efficacy and other reasons not necessarily related to adverse events or side effects. Further, all studies lost participants to attrition. Dropouts from observational studies are more likely to bias reporting toward lack of adverse events. Patients on long-term treatment are self-selected for positive outcomes, even though this may be due to any combination of treatment effects and benign disease course. The studies also did not use large enough patient populations to adequately detect rare events.
Eight observational studies (all high risk of bias) addressed the risks of rebound disease activity with natalizumab treatment interruption. Determining whether rebound exists requires comparing disease activity prior to receiving natalizumab and disease activity after interrupting treatment. None of the studies used appropriate comparison groups. All but one study used a subjective definition for rebound.

Twelve observational studies (all high risk of bias) addressed the benefits and risks to mothers and fetuses of IFN, glatiramer acetate, or natalizumab treatment discontinuation due to pregnancy or intended pregnancy. Women who discontinue DMT with the intention of becoming pregnant risk increased relapses between discontinuation and pregnancy, as well as postpartum. Given that the studied populations are those who became pregnant, none of the studies capture what happens to women who discontinue DMT but do not become pregnant. Therefore, no research has observed whether such women are at increased risk of relapse.

KQ1c. Reasons for Discontinuing DMT

The key points for KQ1c are as follows:

- The broad variation in discontinuation reporting prevented useful aggregation of studies.
- All studies reported one or more adverse events and inefficacy or progression of disability as reasons to discontinue.
- Patient reasons for discontinuing DMT were not explored.

Results are summarized in Table C.

Table C. Studies reporting reasons for discontinuing medication

<table>
<thead>
<tr>
<th>DMT</th>
<th>Total Number of Studies</th>
<th>Adverse Event</th>
<th>Inefficacy or Progression of Disability*</th>
<th>Intended Pregnancy</th>
<th>Long-Term Stable MS</th>
<th>Death</th>
<th>Protocol Violation</th>
<th>Patient Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glatiramer acetate</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Teriflunomide</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Interferon beta-1a</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Interferon beta-1b</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Interferon beta, mixed</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

| DMT, mixed | 3 | 3 | 3 | 3 | 0 | 1 | 0 | 3 |

DMT = disease-modifying treatment; MS = multiple sclerosis

*Category includes counts of discontinuation based both on clinician evaluation of disease progression and patient evaluation of lack of efficacy

Twenty articles of the full reporting set reported reasons for discontinuing treatment. The wide range of reporting methods and discontinuation categories prohibited detailed quantitative aggregation over the studies. Most articles reported numerous reasons for discontinuations. Unfortunately, the reason for the patient’s decision to discontinue remained largely unexplored. Minimal text in this category generally used phrases such as “by own will,”25,37 “withdrew consent,”22,28 or “voluntary withdrawal.”34,38,39
**KQ2. Preferences for Discontinuing DMT**

The 25 included unique studies (28 total articles) represented a wide range of study aims. Designs ranged from factor analysis of questionnaires to experimental psychology lab tests to trials of shared decisionmaking interventions. Study locations were international, including the United States,\(^{40-47}\) the Netherlands,\(^{48-50}\) Germany,\(^{51-61}\) Norway,\(^{62}\) a consortium of European countries,\(^{53}\) Canada,\(^{64,65}\) Italy,\(^{66}\) and Ireland.\(^{57}\)

Given the complexity of understanding preferences and behaviors, and the wide range of study designs used over a small literature set, all KQ2 key points should be viewed as preliminary.

**KQ2a. Intrapersonal Aspects**

The key points for intrapersonal aspects of KQ2 are as follows:

- Patients overestimated intermediate-term risk of wheelchair use but underestimated the lifetime risk. This underestimation may indicate the uncertainty MS patients felt when contemplating their personal trajectories rather than lack of knowledge (2 studies).
- Patients are likely to use heuristics in risk assessments (1 study).
- With training, patients can improve risk understanding and sense of informed choice (1 study).
- Quantified preference studies suggest that patients are willing to make risk tradeoffs for benefits only to the point where the discomfort from side effects and treatment are equal to or worse than the disease symptoms (2 studies).
- Increasing out-of-pocket cost reduces DMT purchases (2 studies).
- Common reasons for discontinuing include side effects, uncertainty about or perceived lack of efficacy against disease progression, administration method and frequency, and cost (5 studies).
- MS patients tended to take responsibility for the decision to discontinue (3 studies), while viewing their neurologist as the driver for decisions regarding choice of DMT (1 study).
- Psychological models of behavior support the presence of rational processes contributing to patient decisionmaking (2 studies).

We found 14 studies in the literature on values and preferences to populate the intrapersonal portion of Figure A addressing intrapersonal factors. The literature tended to examine attitudes and cognition rather than patient knowledge and how that knowledge affected decisions. Studies examined risk expectation, preferences for DMT and treatment tradeoffs, knowledge of cost factors, reasons for using or discontinuing DMT, and theoretical approaches to understanding decisionmaking and behavior processes.

**KQ2a. Interpersonal Aspects**

The key points for interpersonal aspects of KQ2 are as follows:

- MS patients and their physicians can differ significantly in their perceptions of the relative importance of health states and risks (2 studies).
- Physicians and patients must communicate in order to clarify differences in perceptions and preferences (1 study).
Much less literature populates the interpersonal than intrapersonal portion of Figure A. Three studies examined interpersonal concerns, including the knowledge, values, beliefs, and preferences that both the patient and physician bring to a decisionmaking encounter, and also the extent to which this information is shared between the two. Communication issues also are important at the interpersonal level.

### KQ2b. Patient and Provider Preferences for Participation in Shared Decisionmaking

The key points for KQ2b are as follows:

- Different MS patients may bring different information-seeking orientations to shared decisionmaking (1 study).
- Mildly cognitively impaired MS patients show a significantly reduced capacity to understand treatment disclosures, but understanding may be brought back to the level of healthy controls through repetition and recognition cuing (1 study).
- The large majority of people with MS prefer a collaborative or active role in treatment decisions (3 non-U.S. studies).
- Physicians cannot reliably predict patient preferences for an active participation role and may inadvertently pull patients away from their preferred treatment (2 studies).
- Both patient and third-party observers rated physicians as showing limited skill at involving patients in shared decisionmaking (1 study).
- Providing balanced evidence-based information alone is not sufficient to alter decisionmaking processes to help patients achieve their preferred participation role (1 study).

Literature for this KQ subquestion relates to shared decisionmaking for patients and providers. All but one of the identified studies populated the center box in the shared decisionmaking portion of Figure A. Five studies addressed shared decisionmaking from the patient side, four addressed the physician side, and one tested a decision aid to improve shared decisionmaking.

### Discussion

Effective health care relies on the three legs of physicians’ clinical experience, patients’ knowledge of their specific health situations and preferences, and an evidence base. Together, these three components provide the input for medical decisions. In the absence of a clear unambiguous path to follow, patients are best served by shared decisionmaking, which requires clinicians to provide the best available information against which patients can weigh their preferences and risk tolerance.

The decisions around discontinuation of DMT are extremely personal and individual. It is hard to envision ever having enough information to cover all contingencies. Providers and MS patients who have followed a prolonged DMT treatment plan have little information to guide decisions regarding discontinuing DMT. Thus, personal preferences about risks take on more weight.

No literature directly compared continuing versus discontinuing DMT in comparable populations. Only sparse information was available to address one part of the decisionmaking picture faced by providers and patients: long-term benefits and harms. As summarized in Table
D, low-strength evidence showed increased all-cause mortality for patients who started IFNbeta-1b 2 years earlier than the comparators, but no differences between treated and comparator groups in time to progression to SPMS (as measured by a sustained EDSS score of 6). Similarly, overall long-term harms were found to be no different from short-term harms. Low-strength evidence implies low confidence in the findings and the expectation that future research could change the findings. Evidence is insufficient to assess long-term benefits and harms for any other patient population, type of DMT, or outcome.
The current literature did not examine whether long-term benefits for DMTs remain after a patient converts to SPMS. For the special cases of natalizumab and discontinuation for planned pregnancy, evidence was insufficient to answer whether discontinuation is problem free.

In the absence of evidence, providers and patients are left with little to inform their preferences and guide their decisions regarding when to discontinue treatments. The majority of included studies reported reasons for patients discontinuing treatments, but the information provided was without detail. Adverse events and inefficacy or progression of disability were two expected categories. Other possible reasons for discontinuation, such as a patient’s desire to try alternative medicine approaches, perceived risk of long-term use, or financial concerns such as out-of-pocket costs or loss of insurance, are not noted. The “patient decision” category for discontinuing was consistently unexplored.

Harms from long-term DMT use, as is true for many treatments and medical conditions, is poorly reported in the literature. The low-strength evidence showing long-term harms to be generally similar to short-term harms may very well be upset by improved tracking and reporting. For example, a November 2014 FDA Drug Safety Communication reported the first confirmed fatal case of PML for an MS patient using dimethyl fumarate (Tecfidera®). The patient had used dimethyl fumarate for 4 years.

KQ2 aimed to delve into what is known about patient and provider preferences. While the literature was sparse, with only 28 studies available to populate the conceptual map provided in Figure A, each of the three major conceptual areas was at least partially represented. No study directly asked why people are reluctant to discontinue when treatment seems no longer effective, but taken as a whole, the literature set provides some insight.

Overall, one can weave together the general themes found in the KQ2 literature. Admittedly, physicians cannot reliably predict patient preferences for shared decisionmaking, and often physicians and patients perceive the relative importance of health status or acceptable risks differently. However, when it comes to the decision to discontinue DMT, the patient drives the

### Table D. Summary of KQ1 findings with sufficient evidence

<table>
<thead>
<tr>
<th>DMTs Used in Long-Term Studies Assessing Discontinuing or Continuing DMTs</th>
<th>Number of Studies; Number of Participants</th>
<th>Findings</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cause survival: interferon beta-1b</td>
<td>1 study; N = 366 RRMS</td>
<td>All-cause mortality: 250 mg arm vs. placebo—HR, 0.532 (98% CI, 0.31 to 0.90) 50 mg arm vs. placebo—HR, 0.54 (95% CI, 0.32 to 0.92) Favors treatment</td>
<td>Low (moderate risk of bias, unknown consistency)</td>
</tr>
<tr>
<td>Time to progression to SPMS: interferon, mixed</td>
<td>1 study; N = 2,656 RRMS</td>
<td>No difference from contemporary or historical control</td>
<td>Low (moderate risk of bias, unknown consistency)</td>
</tr>
<tr>
<td>Overall harms: interferon, glatiramer acetate, teriflunomide</td>
<td>3 studies; N = 746 RRMS, interferon beta-1b, 16 years; N=46 RRMS, glatiramer acetate, 22 years; N=131 RRMS, 16 SPMS, teriflunomide, 8.5 years</td>
<td>Long-term harms not different from short-term harms (qualitative finding)</td>
<td>Low (high risk of bias, consistent, indeterminate precision)</td>
</tr>
</tbody>
</table>

CI = confidence interval; DMT = disease-modifying treatment; HR = hazard ratio; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis
decision, and this preference and role are generally unchallenged by the physician. In some DMT discontinuations, the balance of shared decisionmaking may shift to discordance between the physician and patient, with the physician deferring to patient preferences for continuing treatment or not. The quantified preferences work by Prosser and colleagues illustrates a paradox—patients are less likely to prefer DMT during the early course of the disease, when disease symptoms are lower than the side effects of the DMT, and more likely to use it at later stages of the disease, when the side effects are less than disease symptoms. This behavior is counter to the hypothesis under which DMTs are assumed to work, which is by using DMTs to reduce relapses early in the disease course to prevent or delay disease progression. Without more solid evidence for the long-term net benefits or the thresholds at which treatment is no longer effective in preventing disease progression, the decision to discontinue treatment remains preference sensitive.

The preferences literature underscores the complexity of the topic and the processes underlying decisionmaking. Both rational and nonrational (such as heuristic) processes came into play, and neither had primacy over the other. Cost was a factor in both self-report and through observation of purchasing behavior. Cognitive deficits impairing decisional capacity may be overcome with adequate cuing. Information is a necessary component of decisionmaking, yet nonrational factors can influence what information is sought at what time.

Preferences, values, and beliefs are highly variable, may change over time, and are linked to the nature of the patients’ relationships with their doctors. There may well be differences based on age, sex, race, class, and other factors. A patient’s preference position between “treat my MS at any cost/comfort from knowledge of receiving treatment” and “need strong evidence that the medication will help and be worth the cost/side effects” may change over time and as the disease changes.

Changing perceptions regarding health states were common across different parts of the intrapersonal literature. Risk perceptions and quantified preferences (which are risk based as well) both suggested that people with longer MS experience assigned higher values to disabled states or viewed them as less serious than people with shorter MS experience did. This is a finding consistent with other research into how people value different health states. Many people overestimate their aversion to hypothetical states of disability and hence eliminate treatment options that might lead to such disability, especially if it could be long term. The hypothetical disutilities for these states are consistently higher than the actual disutilities for those experiencing the state.

Issues
Several challenges impede the gathering of evidence to inform decisions to discontinue DMT. First, the potential differential effectiveness of DMTs for different patient subpopulations is unclear because of the lack of studies examining the questions as well as the use of unsatisfactory study designs. Whether DMTs for CIS patients are effective remains an open question. DMTs may offer little benefit in exchange for side effects and potential harms for patients with a benign MS course. Conversely, it is not known which patients are at risk of worsened disease activity (such as a rebound effect) when DMTs are discontinued, possibly prematurely. We cannot currently predict early or benign disease courses.

Second, the transition from RRMS to SPMS is difficult to ascertain and therefore poses challenges in the decision to discontinue treatment. Clear biomarkers do not exist, and neither do distinct boundaries for the transition. Currently, clinical judgment and EDSS changes or an
EDSS score of 6 or 7 are generally used. Furthermore, some patients with RRMS never transition to a clear secondary progressive phase. Since relapses tend to decrease in frequency with advancing age (being rare after the sixth decade of life and very rare after the seventh decade), the problem arises of determining whether a patient's lack of relapses is due to ongoing DMT or to the natural history of the disease. For example, consider a 75-year-old patient who developed RRMS at age 30, has been taking DMT since 1994, has had no relapses or new MRI lesions since 1996, and has shown no evidence of secondary progression (stable EDSS). Is the lack of relapses due to ongoing DMT use, or has this patient's MS reached the stage where the risk of relapse is passed and there is no ongoing neurologic deterioration beyond what would be expected in normal aging (sometimes referred to as "burned out" MS)? Is it safe to discontinue DMT in such patients? Adequate data to answer this question are not yet available.

This observation leads to the third major challenge: measuring disability. The EDSS is the most commonly used scale in research, in part because it is the longest standing. Because the EDSS is largely driven by mobility assessment, available research is generally silent on potential benefits of DMT other than ambulation, such as upper limb function and cognitive impairments. Other validated measures of health status in MS that incorporate more function domains include the MS Quality of Life-54 (MSQOL-54, the Functional Assessment of Multiple Sclerosis (FAMS), and the Multiple Sclerosis Quality-of-Life Inventory (MSQLI). As seen in KQ2, given that people with MS can value health domains differently than physicians (or perhaps researchers), the broader range of disability assessment should be pursued regardless of any potential limitations in comparing results with studies that used the EDSS exclusively.

Without adequate measures of quality of life, balancing the benefits of treatments against harms becomes challenging, especially across different drug regimens. DMTs are not benign with regard to side effects and risk profiles. The degree to which quality-of-life benefits of treatment are offset by quality-of-life decreases due to side effects and risk profiles is important.

Much remains to be done to understand patient preferences. Emerging but useful information was available to explore KQ2, but no study directly asked the question about preference for discontinuing treatment or explored why patients may be unwilling to discontinue even when treatment no longer appears effective. Lonergan and colleagues approached the question tangentially, asking physicians about how they counsel patients when considering discontinuation. Providers who are involved with such counseling sessions would also benefit from research that separates understanding of preferences, which may be clear to the patient, and the mixed feelings such preferences may generate, ranging from fear or grief related to “giving up” on the disease to relief at no longer carrying the burden of DMTs.

Newly approved drugs, such as fingolimod, and drugs in the development pipeline are emphasizing oral administration to improve medication uptake and adherence to treatment programs. Self-injection can be a deterrent to patients with MS starting injectable DMTs, and “shot-fatigue” is a significant factor for adherence. Oral medications will certainly have implications for preferences for continuing and discontinuing DMTs.

Future Research

Since only three areas of evidence for KQ1 were sufficient to provide answers with only low strength of evidence, essentially all questions related to KQ1 would benefit from further study. The utility of studies for estimating long-term treatment effectiveness in MS can be improved by using prospective population-based designs with appropriate comparators and standardized data collection methods. Study cohorts must be better characterized with respect to demographic and
clinical characteristics, as well as other factors that may influence outcome, such as socioeconomic status, access to care, health behaviors, and comorbidities. Near-complete patient retention with regularly scheduled patient visits is also necessary. The ability to account for treatment effects would improve with better models to predict disability outcomes in MS, including disentangling the young versus old from the new versus long-term disease presence, since the two overlap. Techniques to adjust for selection bias, such as regression analysis or propensity scores, are more easily accomplished with rich datasets. Since the pharmaceutical industry would not benefit from strong comparator studies focusing on treatment discontinuation, other funding sources will need to be identified. With regard to the question of discontinuing for pregnancy, appropriate comparison groups need to include women who discontinued DMT to attempt pregnancy but did not conceive.

Some efforts to improve longer term research are underway—for example, the prospective 5-year OPT-Up study. While the study is more geared toward initial treatment and switching choices, understanding discontinuation within that context is one of its goals. A prospective 10-year observational study based on the United Kingdom’s MS risk-sharing scheme is evaluating the effectiveness of the first DMTs, IFN and glatiramer acetate. After the National Institute for Health and Care Excellence (NICE) recommended against DMTs in 2002, a pricing scheme was negotiated with participating pharmaceutical companies whereby the drug prices would be reduced if patient outcomes were lower than expected; thus, the United Kingdom National Health Service and the pharmaceutical companies shared the financial risk for cost-effective treatment. The initial 2-year results, published in 2009, found that patient outcomes were worse than predicted. However, results were controversial; an independent review of the data identified intrinsic flaws in both the control dataset and analysis model selected when setting up the risk-sharing scheme. Four-, 6-, and 8-year data have been collected and are being analyzed using an updated modeling methodology. This research initiative should help inform the long-term benefits of these injectable treatments and may suggest improvements to current MS registries or methods, making analysis of such registries more fruitful.

KQ2 covered a broad array of relevant topics, and investigator-driven research remains a likely source for innovative and interesting approaches to continued exploration. The AutoMS project, an international consortium of six European locations and Australia, was formed in 2010 to explore MS patient preferences for shared decisionmaking. Confirming the generalizability of their findings to the United States would be beneficial. Also useful would be well-designed qualitative and survey research, perhaps as a mixed-methods study, exploring why and under what circumstances a patient might seek to terminate treatment, and why people are reluctant to discontinue when treatment appears no longer effective.

Attention to areas such as implementation science and quality improvement, which combined are often referred to as health care improvement science, should be included in the future of MS research and the improvement of MS treatment decisions and outcomes. Such areas may contribute systems-level factors to DMT selection and adherence, and to the successful implementation of shared decisionmaking.

**Limitations**

Literature on preferences is not indexed to permit easy identification of relevant articles. Search strategies to capture the diffuse literature used natural language as keywords. While we tested multiple terms before settling on the final algorithm, relevant articles were likely missed, and thus the included literature set must be viewed as comprehensive but not exhaustive.
Likewise, setting the review scope to exclude adherence literature, as adherence by definition connotes a decision to continue DMT use, may have precluded some relevant literature examining lack of adherence as a de facto decision to discontinue use.
References


Introduction

Background

Since the efficacy of DMTs has been recently reported elsewhere, the goal of this report was to address those patients who have used DMT for prolonged periods. The report presents the state of the science that helps inform the decision to continue or discontinue DMTs and the place that shared decisionmaking by patient and physicians has in such situations.

Condition

Multiple sclerosis (MS) is a variably debilitating disease characterized by demyelination (deterioration of the protective myelin sheaths covering nerve cell processes in the brain and spinal cord) and axon loss within the central nervous system. The lesions created by the myelin destruction and resulting scar tissue interfere with normal transmission along nerve fibers within the brain and to and from the brain. This results in classic symptoms associated with MS. The condition affects 2.5 million individuals worldwide and approximately 400,000 in the United States.1 Twice as many women as men are affected, and diagnosis usually occurs between the ages of 20 and 50.1 About 40 percent of people with MS receive some form of disability income.2 Both symptoms and disease course are highly individualized, depending on where the lesions occur within the central nervous system and the type of MS. Clinically definite MS types have been commonly described as:

- Relapsing-remitting MS (RRMS) is the most common form, affecting approximately 85 percent of patients. Patients typically are diagnosed in their 20s or 30s. Neurologic symptoms of a relapse typically develop over a course of days, stabilize, and spontaneously improve; however, over time permanent disability often accrues with further relapses. Many patients with RRMS will eventually transition to secondary progressive MS (below); estimates of the median time to this transition range from 15 to 29 years.3

- Secondary progressive MS (SPMS) is characterized by worsening disability with or without relapses. Patients may have exacerbations, but the trend over time is a relatively steady progression of disease and disability.1

- Primary progressive MS (PPMS) represents about 15 percent of patients and affects women and men about equally. This form has the worst prognosis and is characterized by gradual and progressive worsening of function without distinct relapses.1

- Progressive relapsing MS (PRMS) affects about 5 percent of patients. This form is usually initially diagnosed as PPMS due to a steady worsening of functioning and changed to PRMS when the patient experiences a relapse. Recently, a recommendation has been made to eliminate this type, incorporating such patients in PPMS.4

People with clinically isolated syndrome (CIS), a first neurological episode, consistent with an MS relapse, may or may not go on to develop MS. CIS involves neurological symptoms such as vision loss, numbness, or weakness that lasts at least 24 hours and are caused by inflammation or demyelination in one (monofocal) or more (multifocal) sites in the central nervous system. In a cohort of 107 CIS patients followed for 20 years, 60 patients with three or more lesions (seen via magnetic resonance imaging [MRI]) converted to definite MS, while only seven with normal baseline MRI converted.5
Debate continues to surround the underlying etiology of MS. Most literature considers MS as an autoimmune, inflammatory disease, and this hypothesis underlies current drug treatments. Others suggest that MS is best understood as a neurodegenerative disease, leading to autoimmune reaction to the neurodegenerative debris. Still others hypothesize that MS is a chronic metabolic disorder. Disease-modifying treatments (DMTs) are aimed at modifying autoimmune activity, even though we do not yet know whether autoimmune activity is primary or secondary in MS pathology.

**Treatment Strategies – and Their Discontinuation**

MS cannot be cured with current therapies, and clinicians and patients face challenges balancing benefits and risks when choosing a treatment course. Current DMTs comprise immunomodulating and immunosuppressant medications, which aim to slow the progression of MS and improve quality of life. The working hypothesis is that reducing or preventing new lesions and their sequelae slows the worsening of disease.

Given the unpredictable nature of MS relapses and progression, clinical trials for MS treatments are run for 2 to 3 years to allow for more accurate measurement of reductions in relapse rates and evidence of slowed disease progression. Table 1 provides a list of currently FDA approved DMTs.

### Table 1. FDA-approved disease-modifying treatments for MS

<table>
<thead>
<tr>
<th>Generic (Administration)</th>
<th>Manufacturer (Trade Name)</th>
<th>FDA-Approved Indication</th>
<th>FDA Warnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interferon beta-1a (IFNb-1a) (Injection: Avonex-weekly, Rebif-thrice weekly, Plegridy – every 2 weeks)</td>
<td>Biogen (Avonex&lt;sup&gt;®&lt;/sup&gt;) EMD Serono (Rebif&lt;sup&gt;®&lt;/sup&gt;)</td>
<td>May 17, 1996, for CIS and RRMS August 2014 (Plegridy) March 7, 2002, for RRMS</td>
<td>Yes Yes</td>
</tr>
<tr>
<td>Interferon beta-1b (IFNb-1b) (Injection every other day)</td>
<td>Bayer Healthcare Pharms (Betaseron&lt;sup&gt;®&lt;/sup&gt;) Novartis (Extavia&lt;sup&gt;®&lt;/sup&gt;)</td>
<td>July 23, 1993, for RRMS August 14, 2009, for CIS and RRMS</td>
<td>Yes Yes</td>
</tr>
<tr>
<td>Glatiramer acetate (Injection-daily)</td>
<td>Teva (Copaxone&lt;sup&gt;®&lt;/sup&gt;)</td>
<td>December 20, 1996, for RRMS</td>
<td>Yes</td>
</tr>
<tr>
<td>Mitoxantrone (IV)</td>
<td>Bedford; Hospira; Teva Parenteral; Fresenius; EMD Serono Inc., Kabi USA, Mylan Institutional; Onco Therapies LTD</td>
<td>2000 for RRMS, SPMS, PRMS</td>
<td>Yes, black box</td>
</tr>
<tr>
<td>Natalizumab (IV)</td>
<td>Biogen (Tysabri&lt;sup&gt;®&lt;/sup&gt;)</td>
<td>November 23, 2004, for RRMS</td>
<td>Yes, black box</td>
</tr>
<tr>
<td>Teriflunomide</td>
<td>Sanofi Aventis US (Aubagio) (leflunomide by Sanofi Aventis US as Arava for arthritis)</td>
<td>September 12, 2012, for RRMS</td>
<td>Yes, black box</td>
</tr>
<tr>
<td>Fingolimod (Oral)</td>
<td>Novartis (Gilenya)</td>
<td>September 21, 2010, RRMS</td>
<td>Yes</td>
</tr>
<tr>
<td>Dimethyl fumarate (Oral)</td>
<td>Biogen (Tecfidera)</td>
<td>March 27, 2013, RRMS</td>
<td>Yes</td>
</tr>
<tr>
<td>Alemtuzumab (Injection)</td>
<td>Genzyme (Lemtrada)</td>
<td>November, 2014, RRMS</td>
<td>Yes</td>
</tr>
</tbody>
</table>

CIS = clinically isolated syndrome; FDA = U.S. Food and Drug Administration; IV = intravenous; MS = multiple sclerosis; PRMS=primary relapsing multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis

A 2013 Cochrane overview review and network analysis of 44 trials of DMTs for MS found moderate to high quality evidence that DMTs are effective against recurrence of relapses in RRMS during the first 24 months of treatment compared with placebo. The network analysis ranked natalizumab as the most effective drug, followed in order by IFNb-1a (Rebif), mitoxantrone, glatiramer acetate (Copaxone), and IFNb-1b (Betaseron). Confidence in the
Evidence dropped to moderate for direct comparisons of mitoxantrone and IFN beta-1b versus placebo and very low for glatiramer acetate versus placebo. Included trials did not report relapse outcomes for RRMS at 3 years. Further, natalizumab and IFN beta-1b were more effective than IFN beta-1a in reducing the number of RRMS participants with disease progression as measured with surrogate markers. In patients with progressive MS, both pairwise and network analysis found no DMTs analyzed prevented disability progression over 2 or 3 years. The overview and network analysis were too recent to include the newest approved drugs such as fingolimod or dimethylfumerate.

Unfortunately, the efficacy of MS treatments appears to correlate with the frequency and severity of side effects. The injectable treatments, the interferon drugs and glatiramer acetate, were modestly efficacious and side-effects were tolerable by many patients. Mitoxantrone, an escalation medication, has a lifetime maximum dosage due to cardiotoxicity, and risks of leukemia. Natalizumab, the first monoclonal antibody approved for treating MS, can induce the potentially fatal brain infection progressive multifocal leukoencephalopathy (PML). Risk for PML increases with natalizumab use longer than 2 years, anti-JC virus antibody status, and prior use of immunosuppressive agents. People taking natalizumab may take a drug holiday or discontinue use completely if their risk factor increases, assessed by a positive test for the anti-JCV antibody status.

Women considering pregnancy face special considerations for drug holidays. There are no class A drugs (drugs safe for use during pregnancy, according to the Food and Drug Administration) for MS. Women and their physicians must weigh the possible risks of DMT exposure to the unborn fetus against the maternal risk of disease progression if she discontinues DMT. Rates of fetal exposure to DMTs vary greatly by country. While a Canadian study reported an exposure rate of 5 percent, much higher rates have been reported by studies outside of North America. A Spanish study found 39 percent of pregnancies were exposed to DMTs, and a Brazilian study found an exposure rate of 70 percent. Unfortunately, conception isn’t always easily planned.

The optimal duration of DMT use remains an open and controversial question. Many patients do not use these medications throughout their entire life after diagnosis. However, with few exceptions (such as natalizumab use or intended pregnancy), patients who opt for DMT for MS may end up using it for several years to decades, as long as they tolerate the treatment and the DMT seems effective. Patients may switch between DMTs in order to find one that is more effective or more tolerable, and studies have found high rates of switching between drugs. Some patients cannot tolerate any of the DMTs, but if a tolerable drug regime is determined, treatment generally continues until the individual reaches a disease stage where DMTs are no longer considered to be helping. Such a point may be reached when a person is determined to be nonresponsive to the medication due to disease progression. Determining when DMT is no longer helpful is challenging. Thus, major questions of interest are whether or not DMTs for MS alter the natural history of the disease in the long run and when (if ever) to discontinue DMT.

**Scope and Key Questions**

**Scope of the Review**

This review examines the long-term (more than 3 years) consequences of discontinuing DMT. We looked for evidence that directly assessed discontinuing versus continuing DMT, and also evidence for long-term (more than 3 years) benefits and harms for either continuing or
discontinuing, since that decision can be informed by the benefits or harms directly linked to either course of action. This information would extend understanding beyond the short-term trials examined in the 2013 Cochrane review. We were also interested in the reasons for discontinuing treatment reported in the long-term studies.

We concentrated on outcomes that patients notice or factor directly into their decisionmaking, such as relapse rates and changes in disability level, rather than intermediate outcomes such as lab tests for neutralizing antibodies. MRI to identify MS-related lesions has been shown to correlate with short-term relapse rates, 6 months to 2 years. However, long-term MRI followup as a surrogate marker for relapse rates or, more importantly, disease progression, currently lacks evidence. Further, short-term MRI followup as a predictor of long-term disability progression answers a different research question, that is, does short-term treatment effect the long-term outcomes, than the research questions asked for this review. Thus, we did not use MRI as a long-term outcome in this review. However, we included MRI as a short-term outcome in the subset of patients discontinuing natalizumab due to risk of PML.

People with MS commonly switch between the available DMTs depending on tolerance, presence of adverse effects, and perceived helpfulness of the treatment. The pertinent clinical question for switching medications is how to define the threshold of disease activity for changing medications. This important question is qualitatively different than that of when, if ever, to stop DMTs completely. To adequately address the question of when to switch medication will likely require a review of both short- and long-term research. Therefore, questions related to switching between DMTs are outside the scope of this review.

The review also examines the evidence for patient values, beliefs, and preferences regarding discontinuing DMTs. Such information should support clinicians, patients, consumer advocates, and other decisionmakers in understanding the factors and processes that may inform decisions to discontinue treatment.

**Key Questions**

We synthesized the evidence in the published literature to address the following two Key Questions (KQ):

**KQ1: What are the consequences of discontinuing disease-modifying treatments in adult patients?**

a. What is the evidence for benefits for continuing versus discontinuing treatment?

b. What is the evidence for long-term harms?

c. What reasons for discontinuation of disease-modifying treatments have been reported in long-term observational cohort studies?

**KQ2: What are individual values, beliefs, and preferences regarding discontinuing disease-modifying treatments?**

a. What are patient and provider preferences for discontinuation of disease–modifying treatments?

b. What are patient and provider preferences for participation in shared decisionmaking to discontinue disease-modifying treatments?

Figure 1 provides a conceptual framework that links the KQs. At the top it depicts the logic path both physicians and patients must travel when considering DMTs:
• Does it work?
• What drug should I start with?
• When should I switch a patient to a new drug and what should that drug be?
• When should a patient discontinue disease-modifying treatment?

This logic path describes the context within which patients and clinicians consider clinical factors—tolerability of the medication, disease characteristics at the time of discontinuation (relapses, progression, MRI activity), risk of ongoing disease treatment, other impediments to continued medication use (e.g., difficulty of obtaining, injecting or ingesting, cost, etc.)—and make decisions about DMTs or, in the case of this review, discontinuation (KQ1). The lower part of the figure depicts the progression from an individual's internal decision context and process (such as preferences, values, knowledge, beliefs, and cognitive behaviors and habits) to an interpersonal decision context and processes between the physician and patient. The overlapping ovals representing the clinician and the patient indicate information shared between the two parties versus information and other cognitive processes specific to one individual. Any overlap depends in part on the level of sophistication a patient brings to the decisionmaking process and in part on how well a physician understands a patient’s beliefs, values, goals, and preferences. For example, a patient newly diagnosed with MS in the novice phase of learning about MS would likely have a smaller overlap.19 The interaction between the physician and patient results in decisions that can vary in their level of concordance.
Organization of This Report

The report describes our review methods, including our search strategy, inclusion and exclusion criteria, approach to reviewing abstracts and full publications, and our method for extracting data and summarizing evidence. We also describe the approach to grading the quality of the literature and to evaluating the strength of the body of evidence.

The results section synthesizes the findings by KQ. For KQ1, we report the available evidence on consequences of discontinuing DMT, including long-term benefits and harms for DMTs. We also report the evidence regarding special cases of treatment discontinuation due to pregnancy and natalizumab discontinuation due to increased risk of PML. For KQ2, we present the empirical literature that populates the conceptual framework provided in Figure 1.
Methods

The methods for this review follow the methods established for the Agency for Healthcare Research and Quality (AHRQ) Effective Health Care Program. Key Question (KQ) 1 uses methods suggested in the AHRQ “Methods Guide for Effectiveness and Comparative Effectiveness Reviews” (available at www.effectivehealthcare.ahrq.gov). Since KQ2 was not amenable to usual comparative effectiveness review (CER), we approached KQ2 using Technical Brief methodology. This section focuses on the elements of the protocol; certain methods map to the PRISMA checklist. All methods and analyses were determined a priori.

Topic Refinement and Review Protocol

Initially a panel of key informants comprised of MS researchers, clinicians, consumer advocates, and consumers gave input on the Key Questions (KQs). The draft KQs were then posted for public comment from May 31, 2013, through June 30, 2013, and revised as needed. We then drafted a protocol for the review and recruited a panel of technical experts to provide high-level content and methodological expertise during the development of the review. The Key Informants and members of the Technical Expert Panel (TEP) were required to disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts. Any potential conflicts of interest were balanced or mitigated. Neither key informants nor members of the TEP performed analysis of any kind, nor did any of them contribute to the writing of this report. Members of the TEP were invited to provide feedback on an initial draft of the review protocol, which was then refined based on their input, reviewed by AHRQ, and posted for public access at the AHRQ Effective Health Care Website.

Role of the AHRQ Task Order Officer

The Task Order Officer (TOO) was responsible for overseeing all aspects of this project. The TOO helped to develop a common understanding among all parties involved in the project, resolved questions and ambiguities, and addressed our queries regarding the scope and processes of the project. The TOO reviewed the report for consistency, clarity, and to ensure that it conforms to AHRQ standards.

Analytic Framework

Figure 2 provides an analytic framework describing the treatment path and long-term benefits and harms of continuing versus discontinuing disease-modifying treatment for KQ1. The conceptual framework in Figure 1 provides a conceptual basis for KQ2 literature.
Figure 2. Analytic framework for discontinuing disease-modifying treatments for MS

KQ2: Individual values, beliefs, and preferences for discontinuing DMT

Discontinuing vs. continuing long-term DMT

(KQ1a)

Final health outcomes
- Relapse rate
- Change in disability
- Progression of disease
- Time to sustained disease progression

(KQ1b)

Adverse effects of continuing or discontinuing long-term DMT

Other reasons provided for discontinuing treatment

(KQ1c)

Adult Patients With MS on DMT

DMT = disease-modifying treatment; KQ = Key Question; MS = multiple sclerosis

Literature Search Strategy

Search Strategy

We used bibliographic database searching to identify previous randomized controlled trials (RCTs) and observational studies published from 1990 to August 2014 for studies enrolling adults with clinically isolated syndrome (CIS) or multiple sclerosis (MS). Relevant bibliographic databases for this topic include MEDLINE®, the Cochrane Central Register of Controlled Trials (CENTRAL), PsychInfo®, and Scopus. Our search strategy appears in Appendix A. The search strategy used relevant Medical Subject Headings and natural language terms to find studies on the topic. The concept search was supplemented with filters designed to select experimental designs. Bibliographic database searches were supplemented with backward citation searches of highly relevant systematic reviews.

We conducted additional grey literature searching to identify relevant completed and ongoing studies. Relevant grey literature resources include trial registries and FDA databases. We searched ClinicalTrials.gov, the International Controlled Trials Registry Platform (ICTRP), and Health Services Research Projects in Progress (HSRProj) for observational literature. Scientific information packet (SIP) letters and emails were sent to 11 identified relevant industry stakeholders requesting submission of published and unpublished information on their product(s). Four submissions were subsequently received and reviewed. Grey literature search results were used to identify studies, outcomes, and analyses not reported in the published literature that may further inform findings for KQs. Since we did not find RCTs of sufficient length to be included in this review, we were not able to assess publication and reporting bias using the grey literature.
Inclusion and Exclusion Criteria

We included studies based on the PICOTS (population, intervention, comparator, outcomes, timing, setting) framework outlined in Table 2 and further publication or study characteristics outlined in Table 3.

Table 2. Review PICOTS

<table>
<thead>
<tr>
<th>PICOT</th>
<th>Included</th>
<th>Excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>Adults with CIS, RRMS, SPMS, PPMS, or PRMS using FDA-approved DMTs</td>
<td>Pediatric MS patients</td>
</tr>
<tr>
<td>Interventions used in long-term studies assessing discontinuing or continuing DMTs</td>
<td>DMTs IFNbeta-1a, IFNbeta 1b, glatiramer acetate, natalizumab, teriflunomide, fingolimod</td>
<td>Mitoxantrone – it has a lifetime use limit, so ultimately discontinuing is not a choice. (Dimethyl fumarate has not been approved long enough to generate long-term data. Alemtuzumab was approved too late for inclusion in this review.)</td>
</tr>
<tr>
<td>Comparator groups used in long-term studies assessing discontinuing or continuing DMTs</td>
<td>Patients who received placebo, FDA-approved DMTs, or patients who did not receive any DMT.</td>
<td>DMTs not FDA-approved</td>
</tr>
<tr>
<td>Outcomes and Concepts/Topics of Interest</td>
<td>KQ1</td>
<td>KQ 1</td>
</tr>
<tr>
<td>Patient-centered benefits compared with baseline: reduction in annualized relapse rate (at least one relapse); change in disability, change in EDSS, disease progression (determined by functional assessment); time to sustained disease progression Any reported adverse events Any reported reason for discontinuing treatment</td>
<td>Intermediate outcomes Exception: MRI as intermediate outcome exclusion was relaxed for patients discontinuing natalizumab due to changes in risk of PML, since the time frame was within the short-term 2-3 window.</td>
<td></td>
</tr>
<tr>
<td>KQ2</td>
<td>Individuals' attitudes, values, preferences for discontinuing treatments and health states Perceptions of risk and seriousness of health states Factors and processes patients with MS and clinicians use in shared decisionmaking</td>
<td>Adherence to treatment plan</td>
</tr>
<tr>
<td>Timing</td>
<td>Treatment and followup must be greater than 3 years. Exception: timing was relaxed for: Women considering pregnancy or are pregnant Patients discontinuing natalizumab due to changes in risk of PML</td>
<td></td>
</tr>
<tr>
<td>Setting</td>
<td>Outpatient</td>
<td></td>
</tr>
</tbody>
</table>

CIS = clinically isolated syndrome; DMT = disease-modifying treatment; EDSS = Expanded Disability Status Scale; FDA = U.S. Food and Drug Administration; IV = intravenous; MS = multiple sclerosis; PICOTS = population, intervention, comparator, outcomes, timing, setting; PML = progressive multifocal leukoencephalopathy; PPMS = primary progressive multiple sclerosis; PRMS = progressive relapsing multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis
Table 3. Study inclusion criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Criteria for Inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Enrollment</td>
<td>Studies that enrolled adults with CIS, RRMS, SPMS, PPMS, or PRMS</td>
</tr>
<tr>
<td>Study Design</td>
<td>RCTs, nonrandomized controlled trials, prospective and retrospective cohort studies, case control studies, and case series were included for each population and treatment option.</td>
</tr>
<tr>
<td>Time of Publication</td>
<td>1990 forward. FDA-approved disease-modifying drugs were only available in the U.S. after 1993.</td>
</tr>
<tr>
<td>Study Quality</td>
<td>All studies that met inclusion criteria were screened for eligibility. Studies that did not adequately report study information to allow abstraction of treatment and followup duration or that had indeterminable numerators and denominators for outcomes and adverse event were excluded.</td>
</tr>
<tr>
<td>Language of Publication</td>
<td>English language</td>
</tr>
</tbody>
</table>

CIS = clinically isolated syndrome; FDA = U.S. Food and Drug Administration; PPMS = primary progressive multiple sclerosis; PRMS = progressive relapsing multiple sclerosis; RCT = randomized controlled trial; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis

Study Selection

Bibliographic database search results were exported to EndNote\textsuperscript{21} for screening studies relevant to our PICOTS framework and study-specific criteria. Titles and abstracts were reviewed by two independent investigators to identify studies meeting PICOTS framework and inclusion/exclusion criteria. All studies identified as relevant by either investigator underwent full-text screening. Two independent investigators screened full text to determine if inclusion criteria were met. Differences in screening decisions were resolved by consultation between investigators and, if necessary, a third investigator. We documented the inclusion and exclusion status of citations undergoing full-text screening. A bibliography of studies excluded at full text, and their reasons for exclusion, is provided in Appendix B.

Data Extraction

Studies that met inclusion criteria were distributed among investigators for data abstraction and risk of bias assessment. Two investigators abstracted relevant study, population demographic, and outcomes data. Data fields included author; year of publication; setting, subject inclusion and exclusion criteria; and study design characteristics. For KQ1, we also abstracted intervention and control characteristics (intervention components, timing, frequency, duration); followup duration; participant baseline demographics; type of CIS or MS, MS severity; descriptions and results of outcomes and adverse effects; reasons for discontinuation; and study funding source. Studies that only reported long-term benefits and harms aggregated across multiple disease-modifying treatments (DMTs) were not abstracted. Such studies were accounted for, however, in the article flow-diagram and references are made available. Similarly, studies that did not meet minimum quality levels, or had incomplete reporting to allow analysis, were also accounted for and references made available. Evidence tables are provided in Appendix C.

For KQ2, we abstracted study aims, characteristics, and study findings. Relevant data were extracted directly into summary tables. Summary tables were reviewed and verified for accuracy by a second investigator.

Risk of Bias Assessment of Individual Studies

Because we found only observational literature for this review, we developed an instrument for assessing risk of bias for observational studies using the RTI Observational Studies Risk of Bias and Precision Item Bank for KQ1.\textsuperscript{22} We selected items most relevant in assessing risk of bias for
this topic, including participant selection, ascertainment, attrition, performance, and appropriateness of analytic methods.

Overall summary risk of bias assessments for each individual study were classified as low, moderate, or high based upon the collective risk of bias inherent in each domain and confidence that the results are believable given the study’s limitations. Two investigators independently assessed the studies for risk of bias and consulted to reconcile any discrepancies. When agreement was not reached through consultation, a third party was consulted to reconcile the summary judgment. Information about risk of bias for individual studies is available in Appendix D. Following Technical Brief methods, risk of bias was not assessed for KQ2.

Data Synthesis
For KQ1, we summarized the results into evidence tables and qualitatively synthesized evidence for specific disease-modifying medications and unique population, duration of DMT, length of study followup, and outcomes combinations. Studies were grouped by length of followup to examine changes over time, if any, in outcomes and reasons for discontinuing disease-modifying treatments. We used the best of the available evidence provided by the identified observational literature. Our literature search found no RCTs of sufficient length and significant numbers of observational studies, and many observational studies were found on preliminary examination to have high risk of bias. So while all identified articles underwent abstraction, only the best evidence, based on those studies closest to an “ideal” study design (those studies with the lowest risk of bias) are included in the evidence synthesis.

We synthesized data on several outcomes. The Expanded Disability Status Scale (EDSS) is an ordinal measure assigned by clinical observation, with scores ranging from 0 (no evident disability) to 10 (death). The patient remains at a given score level until enough functional declines are observed to move one step on the scale; for example, from 4.0 to 4.5. A score of 4.5 indicates changes in a person’s mobility. A score of 6.5 indicates a person requires bilateral and generally constant assistance (canes, crutches, braces, walker, or people) to walk 20 meters without resting. Changes in EDSS are commonly used to calculate change in disability or progression of disease. Other measures included changes in the annual or annualized relapse rates from baseline.

For KQ2, we summarized the results into evidence tables and conducted a qualitative synthesis. We grouped the literature by mapping the included studies to the conceptual framework (Figure 1) and analyzed the study findings for emergent patterns for patient perspectives, clinician perspectives, and clinician/patient interpersonal interactions.

Strength of the Body of Evidence
The overall strength of evidence for select outcomes for KQ1 (relapse rate, change in disability, progression of disease, time to sustained disease progression) within each comparison were evaluated based on four required domains: 1) study limitations (internal validity); 2) directness (single, direct link between intervention and outcome); 3) consistency (similarity of effect direction and size); and 4) precision (degree of certainty around an estimate). A fifth domain, reporting bias, was assessed when strength of evidence based upon the first four domains was moderate or high. Based on study design and conduct, risk of bias was rated as low, medium, or high. Consistency was rated as consistent, inconsistent, or unknown/not applicable (e.g., single study). Directness was rated as either direct or indirect. Precision was rated as precise or imprecise. Other factors that may be considered in assessing strength of evidence include dose-response relationship, the presence of...
confounders, and strength of association. Based on these factors, the overall evidence for each outcome was rated as:

- **High**: Very confident that estimate of effect lies close to true effect. Few or no deficiencies in body of evidence, findings believed to be stable.
- **Moderate**: Moderately confident that estimate of effect lies close to true effect. Some deficiencies in body of evidence; findings likely to be stable, but some doubt.
- **Low**: Limited confidence that estimate of effect lies close to true effect; major or numerous deficiencies in body of evidence. Additional evidence necessary before concluding that findings are stable or that estimate of effect is close to true effect.
- **Insufficient**: No evidence, unable to estimate an effect, or no confidence in estimate of effect. No evidence is available or the body of evidence precludes judgment.

Strength of evidence was not assessed for KQ2. This KQ was approached in a hypothesis-generating manner. Following Technical Brief methods, strength of evidence was not assessed for KQ2.

**Applicability**

Applicability of studies was determined according to the PICOTS framework. Study characteristics that may affect applicability include, but are not limited to, type of MS or CIS, unobserved differences in patient preferences, or country within which treatment is provided, given differences in international regulations and treatment preferences.
Results

Results of Literature Searches

We identified 4,359 unique citations searching five bibliography databases (Figure 3) from 1990 to August 2014. After excluding articles at title and abstract, full texts of 198 articles were reviewed to determine final inclusion. Seven articles were added through hand search. Of the 61 articles retained for Key Question (KQ) 1, 11 were specific to discontinuing natalizumab due to increased risk and 12 were specific to discontinuing due to pregnancy. Of the remaining 38 articles comprising 27 unique studies, only 16 studies contained complete information to allow for full analysis. All 38 articles were reviewed for information on reported reasons for discontinuation. For KQ2, 30 articles comprising 27 unique studies were included.

Figure 3. Disposition of studies identified for this review

KQ = Key Question; NTZ = articles for discontinuing natalizumab; PG = articles for discontinuing due to pregnancy
KQ1. Consequences of Discontinuing Disease-Modifying Treatments

Description of Included Studies

Twenty-seven unique observational study populations were represented in 38 articles. (Detailed evidence tables are included in Appendix C.) Studies were conducted in the United States, Canada, Ireland, Italy, Hungary, Spain, Sweden, Denmark, France, Serbia, and multiple international locations. Of these, only 16 studies (24 articles) provided sufficient detail to abstract treatment benefit information. Study comparison arms ranged from early versus delayed treatment, treatment types, treatment dosages, treatment continuers versus treatment stoppers, and treated versus untreated.

Table 4 displays the longest study followup period for each of the 27 studies, although some studies have companion articles that reported specific treatment types or outcome measures with shorter followup periods. The reporting set included all articles meeting all PICOT inclusion criteria. The analytic set used only studies that reported study information with adequate detail to allow abstraction of treatment and followup duration and outcome information. As is often seen in fields working on improving reporting, the articles in the analytic set tend to be newer, and thus tended to report longer followup periods because patient data were available longer. Table 4 presents the analytic set first by type of disease-modifying treatment (DMT), followed by the remaining studies in the full reporting set but not the analytic set.
<table>
<thead>
<tr>
<th>DMT</th>
<th>Author, Year</th>
<th>Duration:</th>
<th>1 Year</th>
<th>2 Years</th>
<th>3 Years</th>
<th>4 Years</th>
<th>5-6 Years</th>
<th>7-9 Years</th>
<th>≥10 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interferons</td>
<td>Portaccio, 2008¹⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.7y β-1a im, 6.0y β-1b</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>β-1a im 3.1y, β-1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carmona, 2008¹⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.6y</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>≤5y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kappos, 2009⁵⁰</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.3y</td>
</tr>
<tr>
<td></td>
<td>Treatment Delayed tx:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment Early tx:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rio, 2007⁷⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>Mean, median treatment duration not stated. 36% stopped treatment during followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bencaik, 2006⁶⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6y</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>6y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Patti, 2006⁸⁸</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6y</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>6y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uitdehaag, 2011¹¹⁻¹³</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6y</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>6y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shirani, 2012⁷⁰</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7-8 y after RCT enrollment</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>96=3 y*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shih, 2012⁴⁸</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cont 3.5y untreated 4.5y</td>
<td>IFNbeta 5.2y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>95=7y*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tedeholm, 2013³⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>RCT 3.3y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bergamaschi, 2012³⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21.1y</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>med. 7.1y</td>
<td>(0.1-8.5)</td>
<td></td>
</tr>
<tr>
<td>Teriflunomides</td>
<td>Confavreux, 2012³⁸</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Goodin, 2012,³⁵,³⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21.1y</td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glatiramer Acetate</td>
<td>Debouverie, 2007³⁷</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.5-8y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>med. 5y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ford, 2010³⁴</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Up to 15y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>withdrawn 4.8y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mill 8.6y ongoing 13.6y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Miller, 2008³⁶⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment Discontinued</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment Discontinued 3.5y(0.7-13y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cont15y(13-22)</td>
</tr>
<tr>
<td>Multiple Drugs</td>
<td>Rio, 2005 and 2006¹¹⁻¹³</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tedeholm, 2013³⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12 y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>Treated all DMT (interferons and glatiramer acetate) as single</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bergamaschi, 2012³⁶</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td>Treated all DMT as single class</td>
<td>up to 35y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*If treatment duration not stated. 36% stopped treatment during followup.
<table>
<thead>
<tr>
<th>DMT</th>
<th>Author, Year</th>
<th>Duration:</th>
<th>1 Year</th>
<th>2 Years</th>
<th>3 Years</th>
<th>4 Years</th>
<th>5-6 Years</th>
<th>7-9 Years</th>
<th>≥10 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interferons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in Analytic Set</td>
<td></td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cunningham, 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>O’Rourke, 2005</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trojano, 2005</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trojano, 2007</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trojano, 2009</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bermel, 2010 and 2013</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Medications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not in Analytic Set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sorensen, 2006</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Milanese, 2005</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Evans, 2012</td>
<td>Followup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Mean unless otherwise noted.

DMT = disease-modifying treatment; IFNbeta = interferon beta; GA = glatiramer acetate; im = intramuscular; med=median; mITT = modified intention-to-treat; MS = multiple sclerosis; NR = not reported; RCT = randomized controlled trial; sc = subcutaneous; Tx = treatment; y = year
KQ1a. Evidence for Benefits for Continuing Treatment Versus Discontinuing

No studies directly assessed continuing versus discontinuing DMT in comparable populations. We therefore turned to literature examining benefits for continuing DMT long-term.

Sixteen observational studies from 24 articles on long-term benefits for DMT provided sufficient detail to abstract treatment benefit information.16,30-36,38,40,44,48-50,52-54,56,57,59-63 (Detailed evidence table are included in Appendix C.) Four studies were funded by industry,30,35,59,60 three by governmental funding,40,49,56 three reported not receiving funding,44,48,62 and six did not report funding sources.16,38,50,53,57,64

Table 5 provides outcomes reported by type of DMT and the range of followup.

Key Points

- Low-strength evidence from one moderate risk of bias study suggests long-term all-cause survival is higher for treatment naïve relapsing-remitting MS (RRMS) patients who did not delay starting interferon beta 1b by 2 years and used DMTs for a longer duration than those who started later.
- Low-strength evidence from one moderate risk of bias study suggests that interferon use did not change disability progression for RRMS patients.
- Insufficient evidence was available to assess long-term benefits for DMTs for secondary progressive MS (SPMS) patients, and most outcomes for RRMS patients. Except for those noted above, studies were high risk of bias, had small sample sizes, and reported effects were small in magnitude.
<table>
<thead>
<tr>
<th>DMT</th>
<th>Author, Year</th>
<th>Type of MS at Baseline</th>
<th>Median or Mean Years to Final Assessment</th>
<th>All-Cause Mortality; Convert to SPMS</th>
<th>Progression</th>
<th>Annual Relapse Rate (ARR)</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFB1a</td>
<td>Uitdehaag, 2011</td>
<td>RRMS</td>
<td>7-8</td>
<td>N (%) convert to SPMS</td>
<td>N (%) with EDSS progression; time to sustained 1-point EDSS progression</td>
<td>N (%) relapse-free; change in ARR from baseline, Annual relapse count by year;</td>
<td>Insufficient</td>
</tr>
<tr>
<td>INFB1b</td>
<td>Kappos, 2009</td>
<td>CIS</td>
<td>5</td>
<td>Primary: time to clinically-definite MS</td>
<td>Change in EDSS from baseline</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carmona, 2008</td>
<td>RRMS</td>
<td>5.6</td>
<td>N (%) convert to SPMS</td>
<td>Change in EDSS from baseline; time to 1-point EDSS progression</td>
<td>Change in ARR from baseline</td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>Bencsik, 2006</td>
<td>RRMS, RPMS</td>
<td>6</td>
<td>Change in EDSS from baseline</td>
<td>Change in EDSS from baseline</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rio, 2007</td>
<td>SPMS</td>
<td>5</td>
<td>Time to 1-point increase in EDSS (or 0.5 points if above 6 at entry)</td>
<td>Change in ARR, % decrease in relapse rate</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Goodin, 2012</td>
<td>RRMS</td>
<td>21</td>
<td>All-cause mortality: 250 mg arm vs. placebo – HR 0.532 (98% CI, 0.31-0.90) 50 mg arm vs. placebo – HR 0.54 (95% CI, 0.32-0.92) favors treatment</td>
<td>Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5. Outcomes reported from unique studies included in the analytic set for long-term DMT use**
<table>
<thead>
<tr>
<th>DMT</th>
<th>Author, Year</th>
<th>Type of MS at Baseline</th>
<th>Median or Mean Years to Final Assessment</th>
<th>All-Cause Mortality; Convert to SPMS</th>
<th>Progression</th>
<th>Annual Relapse Rate (ARR)</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFB Mixed</td>
<td>Portaccio, 2008</td>
<td>RRMS</td>
<td>4.2</td>
<td></td>
<td></td>
<td></td>
<td>(Discontinuation or harms only reported)</td>
</tr>
<tr>
<td></td>
<td>Patti, 2006</td>
<td>RRMS</td>
<td>≤6</td>
<td></td>
<td></td>
<td></td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>Shirani, 2012</td>
<td>RRMS</td>
<td>4.5-10.5</td>
<td>Time to sustained EDSS ≥6, No difference from contemporary control (HR 1.30; 95% CI: 0.92–1.83) or historical control (HR 0.77; 95% CI: 0.58–1.02)</td>
<td>Change in EDSS from baseline; progression by year to 6 years</td>
<td>Change in ARR from baseline</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Glatiramer Acetate</td>
<td>Deboeverie, 2007</td>
<td>RRMS</td>
<td>3-8</td>
<td></td>
<td></td>
<td>Change in EDSS from baseline; % progression by at least 1-point EDSS</td>
<td>Change in ARR from baseline; mean ARR by year</td>
</tr>
<tr>
<td></td>
<td>Ford, 2010</td>
<td>RRMS</td>
<td>4.8-13.6</td>
<td>Proportion reaching EDSS 4, 6, 8; progression to SPMS</td>
<td>Change in EDSS from baseline</td>
<td>Change in ARR from baseline</td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>Miller, 2008</td>
<td>RRMS</td>
<td>3.5-15</td>
<td>% changed from EDSS&lt;4 to&gt;4; EDSS &lt;6 to &gt;6</td>
<td>Change in EDSS from baseline</td>
<td>ARR over time</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Teriflunomide</td>
<td>Confavreux, 2012</td>
<td>RRMS, SPMS</td>
<td>7.1</td>
<td></td>
<td></td>
<td></td>
<td>Insufficient</td>
</tr>
</tbody>
</table>
Table 5. Outcomes reported from unique studies included in the analytic set for long-term DMT use (continued)

<table>
<thead>
<tr>
<th>DMT</th>
<th>Author, Year</th>
<th>Type of MS at Baseline</th>
<th>Median or Mean Years to Final Assessment</th>
<th>All-Cause Mortality; Convert to SPMS</th>
<th>Progression</th>
<th>Annual Relapse Rate (ARR)</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Drugs</td>
<td>Rio, 2005&lt;sup&gt;51,53&lt;/sup&gt;</td>
<td>RRMS</td>
<td>5</td>
<td>Proportion relapse-free; proportion reaching EDSS of 6</td>
<td>% sustained disability progression</td>
<td>Change in ARR from baseline; proportion with decrease in ARR</td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>Tedeholm, 2013&lt;sup&gt;26&lt;/sup&gt;</td>
<td>RRMS</td>
<td>12</td>
<td>Time to SPMS</td>
<td></td>
<td></td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>Bergamaschi, 2012&lt;sup&gt;46&lt;/sup&gt;</td>
<td>RRMS</td>
<td>20</td>
<td>Proportion converting to SPMS</td>
<td></td>
<td></td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

CIS = clinically isolated syndrome; DMT = disease-modifying treatment; EDSS = extended disability scale score; FU = followup; IFNβ = interferon beta; GA = glatiramer acetate; LTFU = long-term followup; med = median; MS = multiple sclerosis; NR = not reported; RRMS = relapsing remitting multiple sclerosis; SD = standard deviation; SPMS = secondary-progressive multiple sclerosis; y = year
Detailed Synthesis

Variation among the included studies on patient populations, interventions, outcome measurements, and time frames, precluded meaningful pooling. Risk of bias for all but two studies was also high. Selection bias, attrition, and lack of accounting for patient populations over time accounted for the greatest threat to risk of bias. Only two studies provide a low strength of evidence for two benefit outcomes measured long-term for interferons.

One moderate risk of bias study examined all-cause mortality over a 21-year period for 366 patients who had enrolled in a randomized controlled trial (RCT) (98.4% of the original RCT participants) testing interferon beta-1b for treatment naïve RRMS patients in 11 clinics in North America.\(^3^5\) The study’s strength lies in the nearly complete followup of patients and the objective outcome measure. The missing patients were evenly spread across the study arms at two each. After the 2-year RCT, participants used DMTs according to patient and physician discretion, but analysis was based on RCT intention to treat. Participants in the treatment arms showed lower all-cause mortality compared with the placebo arm, with the 250 mg arm slightly lower than the 50 mg arm. The survival rate for the placebo arm was consistent with survival rates reported in MS natural history studies. Median treatment duration for the three groups ranged between 7 years for the placebo group, 14 years for the 50 mg arm, and 12 years for the 250 mg arm. Patients assigned to placebo had both later starts and shorter exposure to DMTs. Thus, the study cannot distinguish between the effects of early use and the effects of long-term use.

Three studies examined the effects of DMTs on disease progression measured by conversion to SPMS.\(^4^0,4^8,5^6\) While the studies were not pooled due to heterogeneity of treatments and methodological approaches, they represent on-target studies that directly attempt to evaluate the effects of DMTs on the long-term patient goal of delaying, or possibly preventing, conversion to SPMS, a more severe disease state. Thus, while only one study provided at least low strength of evidence findings, all three are presented in more detail.

One moderate risk of bias study examined the association between interferon beta use and progression to sustained Expanded Disability Status Scale (EDSS) for 6 of 2656 RRMS patients in Canada.\(^4^0\) Three arms were used: a treatment cohort followed for 5.1 years (interquartile range, 3 to 7 years), a contemporary cohort followed 4 years (interquartile range, 2.1 to 6.4 years), and a historical cohort (drawn from pre-interferon period) followed 10.8 years (interquartile range, 6.3 to 14.7 years). The strength of the study by Shirani et al. lies in the almost complete capture of MS patients since patients were unable to obtain DMTs other than from the participating clinic, and the multiple statistical approaches used to test for association, including use of comorbidities (Charlson score) and socioeconomic status along with age, sex, disease duration, and EDSS. Propensity score adjustments did not substantially change the results. The study did not find statistically significant differences in hazard rates for reaching EDSS 6 for either contemporary or historical cohort comparisons. Analyses used a time-dependent treatment variable and accounted for changing treatment status over time to address immortal time bias, a special form of selection bias. The study does not report the percent of patients prescribed different DMTs, or patients with neutralizing antibodies.

One high risk of bias study used selected patients’ data from a Swedish MS registry, also using both a contemporary and historical cohort as comparisons.\(^5^6\) Adjusting for sex, age at MS symptom onset, whether onset was monofocal or polyfocal, and location of lesions, the study found a statistically significant difference in time to SPMS conversion for DMT (interferons and glatiramer acetate) treated patients versus a historical control of untreated patients. Treatment
initiation (whether treatment was delayed or began earlier in the disease course) did not rise to statistical significance. Treatment duration was not a factor in the model. The reported results are inconsistent with Shirani et al. Significant concerns regarding selection bias of the treatment and contemporary groups and treatment by indication determined the study’s high risk of bias.

A second study published only as an e-publication ahead of print used Bayesian techniques to model propensity for treatment – Bayesian risk estimate for MS (BREMS) – and then Bayesian modeling to estimate effects for DMT use in aggregate. The study reports that the technique they used allowed for typical switching and start/stop usage. However, there is insufficient information to assess model quality, and guidelines for model risk of bias are still new to the field and have not yet been translated to systematic review use; thus the risk of bias remains unclear. Bergamaschi et al. found that a smaller proportion of patients using DMTs (the vast majority of patients used interferons or glatiramer acetate) converted to SPMS compared to the historical control. This result is inconsistent with the other two studies.

Insufficient evidence exists to address long-term benefits for glatiramer acetate, teriflunomide, and natalizumab for either RRMS or SPMS, as well as other important MS outcomes for interferon beta for RRMS beyond all-cause mortality or 5-year disability progression for interferon beta.

KQ1b. Evidence for Long-Term Harms

Eleven of the 16 unique studies reported harms in enough detail for abstraction. (Detailed evidence table is included in Appendix C.) Only one of the studies was moderate risk of bias; all others were rated as high risk of bias.

Table 6 provides categories of harms reported by type of DMT and the range of followup.

Key Points

- Limited low-strength evidence suggests harms for injectable DMTs do not differ between short term (2-3 years) and long term (up to 16 years for interferon, 22 years for glatiramer acetate, and 8.5 years for teriflunomide).
- The majority of discontinuation tends to occur in the short-term (2 to 3 years from start).
- Broad variation in harms reporting precludes informative aggregation and summary.
<table>
<thead>
<tr>
<th>DMT</th>
<th>Number of Studies</th>
<th>Total N</th>
<th>Followup</th>
<th>Any Adverse Event</th>
<th>At Least One Serious Adverse Event</th>
<th>Treatment Discontinuation for Adverse Event</th>
<th>Comparator Groups</th>
<th>Reported Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNbeta-1a&lt;sup&gt;16&lt;/sup&gt;</td>
<td>1</td>
<td>N=429</td>
<td>4 y.</td>
<td>Most common AEs: Injection site reactions, headache, flu-like symptoms</td>
<td>NR</td>
<td>NR</td>
<td>No</td>
<td>No difference from short-term events</td>
</tr>
<tr>
<td>IFNbeta-1b&lt;sup&gt;36,46&lt;/sup&gt;</td>
<td>2</td>
<td>N=746</td>
<td>5 – 16 y</td>
<td>Most common AEs: Injection site reactions, depression, flu-like symptoms, headache</td>
<td>21% to 24%</td>
<td>Discontinuation rates “high” but numbers not reported</td>
<td>No</td>
<td>No difference from short-term events Frequency declined over 16 y time in continuers</td>
</tr>
<tr>
<td>IFNbeta mixed&lt;sup&gt;16,44,53&lt;/sup&gt;</td>
<td>3</td>
<td>N=587</td>
<td>4 - 8 y</td>
<td>Most common AEs: Injection site reactions, depression, flu-like symptoms, headache</td>
<td>NR</td>
<td>3% during LT FU. Discontinue for serious AE more likely to happen early in treatment course (1 y)</td>
<td>No</td>
<td>Headache more likely for β-1a, injection site reactions for β-1b, no other differences between type of IFNbeta Majority of discontinuation occur early/short term.</td>
</tr>
<tr>
<td>INFbeta mixed&lt;sup&gt;34&lt;/sup&gt;</td>
<td>1</td>
<td>N=146</td>
<td>5 y</td>
<td>NR</td>
<td>NR</td>
<td>3.4%, although timing isn’t clear</td>
<td>No</td>
<td>Majority of discontinuation occur early/short term.</td>
</tr>
<tr>
<td>Glatiramer acetate&lt;sup&gt;30,38,57&lt;/sup&gt;</td>
<td>3</td>
<td>N=483</td>
<td>4y – 22 y</td>
<td>Only one reported overall rates: 87.3% Most common AE: injection site reactions</td>
<td>NR</td>
<td>Only one reported overall rates: 4.9% in long-term extension</td>
<td>No</td>
<td>Majority of discontinuation occur early/short term No difference from short-term events</td>
</tr>
<tr>
<td>Teriflunomide&lt;sup&gt;39&lt;/sup&gt;</td>
<td>1</td>
<td>N=147</td>
<td>8.5 y followup</td>
<td>98% of 7 mg dose and 100% of 14 mg dose experienced treatment emergent AE</td>
<td>36% of 7 mg dose and 29% of 14 mg dose</td>
<td>13.6% of 7 mg dose and 13.6% of 14 mg dose for AE</td>
<td>One comparison to general population rates for cancer</td>
<td>No difference from short-term events</td>
</tr>
</tbody>
</table>

AEs = adverse events; DMT = disease-modifying treatment; FU = followup; IFNbeta = interferon beta; GA = glatiramer acetate; im = intramuscular; LTFU = long-term followup; med = median; mg = milligram; MS = multiple sclerosis; NR = not reported; RRMS = relapsing remitting multiple sclerosis; sc = subcutaneous; SD = standard deviation; SPMS = secondary-progressive multiple sclerosis; y = year
**Detailed Synthesis**

The included studies used a wide range of reporting methods and adverse event categories that precluded simple aggregation over the studies. The most commonly reported adverse events were injection site reactions, flu-like symptoms, depression, and headache. Serious adverse events were generally not reported, although two studies gave rates of about 25 percent to 30 percent of participants. Discontinuations tended to occur during the first or second year of the study. When reported, discontinuation rates during long-term followup were low, about 3 to 4 percent. However, the studies generally did not separate discontinuation rates due to adverse events from total discontinuation rates, which would also include perceived lack of efficacy and other reasons not necessarily related to adverse events or side effects. Further, all studies lost participants to attrition. Dropouts from observational studies are more likely to bias reporting towards lack of adverse events. Patients on long-term treatment are self-selected for positive outcomes, even though this may be due to any combination of treatment effects and benign disease course. The studies also did not use large enough patient populations to adequately detect rare events.

The Cochrane overview of reviews of RCTs by Filippini et al. provides another basis for comparing short-term and long-term harms. They assessed treatment discontinuation for any reason as acceptability of treatment. The included RCTs were generally 2 to 3 years followup. They assessed treatment harms as counts of participants with 1) at least one adverse event, 2) at least one serious adverse event as defined by the authors of the primary study, 3) withdrawal due to adverse event at any time during the followup period, 4) serious infections as defined by the authors of the primary study, and 5) a new diagnosis of leukemia, lymphoma, or any other type of cancer during the followup period. They found, similar to our results, that studies tended to report number of adverse events rather than number of participants reporting adverse events, and that the range of adverse events and reporting methods was too diverse to allow for data aggregation. Filippini et al. found no statistically significant differences in serious adverse events, serious infections, or cancer between treatment and placebo groups. However, participants in treatment groups were more likely to withdraw from the studies than placebo groups (OR 2.41, 95% CI, 1.92 to 3.03).

**Natalizumab Discontinuation or Drug Holiday**

Eleven observational studies (one moderate and 10 high risk of bias) addressed the risks of rebound disease activity, disease activity above-and-beyond that experienced by the patient prior to starting natalizumab, with natalizumab treatment interruption. (Detailed evidence tables are included in Appendix C.) Studies were conducted in the United States, Canada, Germany, France, Italy, Denmark, Australia, and Spain. Three studies were funded by industry; one was funded by a combination of government, nongovernmental, and industry; two were unfunded; and five did not report funding sources.

Eight of the 11 included annualized relapse rate and MRI measures to assess rebound, while three included MRI measures only. Reported reasons for natalizumab interruption were drug holiday in five studies, mostly drug holiday or not clearly reported in three studies, and mixed reasons in three studies. Five of the eight studies included RRMS patients only, while the others were mixed or unspecified. Sample size ranged from 13 to 48, with three exceptions: Sorensen (n=375), Jokubaitis (n=536), and O’Connor (n=949).
Mean duration of natalizumab treatment ranged from 12 to 32.4 months, while treatment interruption ranged from median 2.6 to mean 11.3 months. Other DMT use during the natalizumab interruption was heterogeneous. DMT was not administered during the treatment interruption in three studies, glatiramer acetate was used in two studies, fingolimod in three studies, mixed DMTs in three studies, and pulsed steroids in one study, although only two studies administered DMTs to all MS patients. All or most patients used DMTs prior to natalizumab, with details not reported or not analyzed. Three studies did not report prior DMT use, one of which compared patients who switched from natalizumab to fingolimod with cohorts who switched to fingolimod from injectable DMT or no treatment.

Table 7 provides definitions of rebound reported in the studies.

**Key Points**

- Overall, evidence is insufficient for whether the rebound phenomenon exists due to the high risk of bias and small study sample sizes.
- Operational definition of rebound lacks consensus. Rebound is subjectively measured and not clearly defined in studies.
- Evidence is lacking for rebound based on the best available evidence of one large study with high risk of bias.
- Studies that reported cases of rebound had high risk of bias, mostly small sample size, were mixed in providing DMTs during the interruption, and did not account for DMT use prior to natalizumab.

**Table 7. Natalizumab rebound definitions**

<table>
<thead>
<tr>
<th>Author, Year Location Risk of Bias</th>
<th>Reported Rebound</th>
<th>Definition of Rebound/Severe Flare</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARR and MRI Measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jokubaitis, 2014 Australia and International Moderate</td>
<td>None</td>
<td>Quantified: compared ARR pre- and post-natalizumab; also compared ARR and distribution of severe relapse among those who switched to fingolimod from natalizumab, injectable, or no DMT</td>
</tr>
<tr>
<td>O'Connor, 2011 Canada High</td>
<td>None</td>
<td>Quantified: comparison to placebo: peak mean ARR during treatment interruption was not higher than the placebo group</td>
</tr>
<tr>
<td>Havla, 2013 Germany High</td>
<td>3/13</td>
<td>Quantified: Sustained EDSS worsening (&gt;1 EDSS steps) and widespread disease activity (&gt;1 GELs) on MRI</td>
</tr>
<tr>
<td>Havla, 2011 Germany High</td>
<td>4/13</td>
<td>Subjective: severe relapse with sustained EDSS worsening and widespread disease activity on MRI</td>
</tr>
<tr>
<td>Kaufman, 2011 United States High</td>
<td>None</td>
<td>Subjective: severe relapse as has been reported by others subjective: clinically severe flare and mean 16 GELs (range 6-40))</td>
</tr>
<tr>
<td>Kerbrat, 2011 France High</td>
<td>4/27</td>
<td>Subjective: severe relapse and 20+ GELs (range 23-50)</td>
</tr>
</tbody>
</table>
Table 7. Natalizumab rebound definitions (continued)

<table>
<thead>
<tr>
<th>MRI Measures Only</th>
<th>Author, Year Location Risk of Bias</th>
<th>Reported Rebound</th>
<th>Definition of Rebound/Severe Flare</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rossi, 2013 Italy High High</td>
<td>None</td>
<td>Subjective: MRI parameters were not consistent with rebound of disease activity</td>
</tr>
<tr>
<td></td>
<td>Sorensen, 2014 Denmark, High</td>
<td>83/375 or 42/375a</td>
<td>Quantified: Higher individual relapse rate after cessation of NTZ than before NTZ</td>
</tr>
<tr>
<td></td>
<td>Borriello, 2012 Spain High</td>
<td>2</td>
<td>Subjective: severe disabling relapse and a large number of GELs</td>
</tr>
<tr>
<td></td>
<td>Borriello, 2011 Spain High</td>
<td>None</td>
<td>Subjective: disease activity worsens beyond pre-treatment severity</td>
</tr>
<tr>
<td></td>
<td>Miravalle, 2011 United States High</td>
<td>1</td>
<td>Subjective: NR for observational study design</td>
</tr>
</tbody>
</table>

ARR = annualized relapse rate; EDSS = Expanded Disability Status Scale; GELs = gadolinium-enhancing lesions; MRI = magnetic resonance imaging; NR = not reported

4Excluding 51 patients treated with mitoxantrone prior to natalizumab

**Detailed Synthesis**

Determining whether rebound exists requires comparing disease activity prior to receiving natalizumab and disease activity after interrupting treatment. The comparisons need to examine people who discontinued natalizumab due to risk of progressive multifocal leukoencephalopathy (PML) rather than lack of treatment efficacy, compare pre and post measures on an individual level if rebound is heterogeneous, and account for other medications used both prior to natalizumab and after treatment interruption. None of the studies met these criteria.

One study addressed the issue of prior treatment by comparing patients who switched from natalizumab to fingolimod with patients who started fingolimod after interferon beta, glatiramer acetate or treatment naïve. Although this study did not examine individual level outcomes or treatment prior to natalizumab, mean annual relapse rate (ARR) did not increase pre- and post-natalizumab among the 89 patients who used natalizumab followed by fingolimod. Additionally, comparison of quarterly relapse rates for each 3-month period in the 15 months before and 9 months after fingolimod initiation found no differences between the natalizumab to fingolimod and injectable DMTs to fingolimod cohorts. No differences were found between any of the three cohorts except that during the 3-6 month period after beginning fingolimod, relapse rates were significantly lower among the treatment naïve cohort (with only one relapse across the cohort) compared with the post-natalizumab cohort. Relapse severity was assessed by comparing relapse treatment (ambulatory, hospitalization, or none) across the cohorts. Fifteen percent of relapses required hospitalization, and these relatively severe relapses were distributed evenly across the three cohorts and across the full observation period, with no evidence of rebound observed during the transition from natalizumab to fingolimod.

One large (n=949) industry-funded high risk-of-bias study examined 8 months of treatment interruptions gathered from large clinical trials. About 13 percent of patients used other DMTs during the interruption. Peak mean ARR following natalizumab discontinuation remained below the on-study placebo ARR. However, ARR and magnetic resonance imaging (MRI) outcomes
were assessed only as averages, with no examination of pre and post measures for individual patients. Sampling of MRI outcomes was limited. Of the 949 study participants who underwent natalizumab treatment interruption, 339 had baseline MRI measures, and followup MRI during treatment interruption ranged from a sample of 147 at 2-3 months after discontinuation to 60 participants at 6 or more months after discontinuation, the time period with peak MRI activity.

Three small, high risk-of-bias studies also reported no rebound. One study examined only mean ARR, with no MRI outcomes or individual pre/post ARR comparisons in 48 patients, some of which used other DMTs during the interruption.\textsuperscript{65} One study of patients who did not use DMTs during the interruption did not report ARR outcomes after treatment interruption and included a sample of less than 20 with complete MRI data.\textsuperscript{75} One industry-funded study reported that the mean number of gadolinium-enhancing lesions (GELs) was significantly higher prior to natalizumab treatment compared with after natalizumab discontinuation in 40 patients. However, all patients were switched to glatiramer acetate, to which they had been treatment naïve, and thus the study lacked a true comparison group.\textsuperscript{71}

In contrast, six high risk of bias (five small) studies reported cases of rebound disease activity. Four studies used subjective rebound definitions.\textsuperscript{66,68,70,74} None of the studies accounted for prior use of DMTs and all lacked the quantitative detail necessary to generate confidence in the study findings.\textsuperscript{66,68-70,72,74} Two studies each reported four rebounds and one reported three rebounds despite small sample sizes (n=13, 27, 36) but lacked quantitative detail on reported MRI outcomes.\textsuperscript{68-70} Prior to natalizumab, all patients in both studies used some type of DMT (details not reported). During the interruption, patients in one study used no DMTs.\textsuperscript{70} In the other two studies, some patients used DMTs while others did not, with all of the reported rebounds occurring in the group taking no DMTs during the interruption.\textsuperscript{68,69} In all four studies reporting ARR, mean ARR did not increase after natalizumab interruption compared with prior to natalizumab treatment. Two other studies did not report ARR after natalizumab interruption.\textsuperscript{66,74} One found no median change in MRI outcomes and reported two cases of rebound despite pulsed steroids use in all patients.\textsuperscript{74} DMT use prior to natalizumab was not reported; 18 of 23 patients had used mitoxantrone. One study found significantly higher mean GELs among the 12 patients with GELs after natalizumab interruption and described one case of rebound.\textsuperscript{66} All patients used DMTs prior to natalizumab, with details not reported, and none during the interruption. One large study reported 83 rebounds based on individual-level comparison of ARR pre- and post-natalizumab. However, this study did not account for the sample’s high disease activity at baseline, included individuals who stopped natalizumab due to treatment failure, and did not address prior DMT (except to report 42 rebounds after excluding patients who had been on mitoxantrone prior to natalizumab).

Since pre-treatment disease activity may return following natalizumab interruption, comparisons that do not account for DMTs used prior to natalizumab are insufficient. Studies found some evidence that those with higher disease activity prior to natalizumab were more likely to relapse after natalizumab interruption. In one study, those with three or more relapses in the year prior to natalizumab treatment were twice as likely to experience relapse following natalizumab interruption (p=0.04) compared with those with fewer relapses prior to natalizumab.\textsuperscript{71} In this study, 88 percent of the total sample (n=40) were on interferon beta prior to natalizumab, and all received glatiramer acetate during the treatment interruption. Two studies reported trends toward higher ARR prior to natalizumab among those who relapsed during interruption.\textsuperscript{66,68} One study found that relapse on fingolimod was predicted by the number of relapses during the 6 months prior to fingolimod, regardless of whether patients started
fingolimod after natalizumab discontinuation or after injectable treatment or treatment naïve, and among the natalizumab to fingolimod cohort, relapses on fingolimod were not predicted by relapse rates prior to natalizumab.\textsuperscript{73}

### Pregnancy Drug Holiday

Twelve observational studies (all high risk of bias) addressed the benefits and risks to mothers and fetuses of treatment discontinuation due to pregnancy or intended pregnancy. (Detailed evidence tables are included in Appendix C.) Published from 2005 to 2012, these 12 studies comprised 11 cohort studies comparing DMT-exposed to unexposed pregnancies, and one large case series, and ranged in size from 14 to 425 DMT exposed pregnancies. The case series was drawn from a global drug safety database, included 425 pregnancies exposed to IFNbeta-1a with prospectively recorded outcomes.\textsuperscript{77} Studies were conducted in Sweden,\textsuperscript{77,78} Italy,\textsuperscript{79-81} Canada,\textsuperscript{13,82} Brazil,\textsuperscript{15} and Germany,\textsuperscript{83-85} and one was an international study with data from the United Kingdom, Brazil, Argentina, and Mexico.\textsuperscript{86} Two studies were funded by industry,\textsuperscript{77,78} two by governmental organizations,\textsuperscript{13,85} three reported that no specific funding was received for the study,\textsuperscript{83,84,86} and five did not report funding sources.\textsuperscript{14,15,79-82}

Definitions of exposed versus unexposed differed (Table 8). Of the 12 studies, 10 reported on IFNbeta (14-88 exposed pregnancies),\textsuperscript{13,15,77-80,82,83,85,86} six reported on GA (6-41 exposed pregnancies),\textsuperscript{13,15,81,84-86} and one reported on natalizumab (35 exposed pregnancies).\textsuperscript{83} Of the 10 studies reporting results for IFN exposure, 2 studies reported on IFNbeta-1a,\textsuperscript{77,78} and 8 studies did not report results by specific IFNbeta preparation.

#### Table 8. Studies reporting pregnancy outcomes by drug

<table>
<thead>
<tr>
<th>DMT</th>
<th>Total Number of Studies</th>
<th>Total N</th>
<th>Spontaneous Abortion</th>
<th>Fetal Death</th>
<th>Preterm Delivery</th>
<th>Postpartum Relapse Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNbeta1a (sc or im)</td>
<td>2</td>
<td>N=466</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>IFN mixed</td>
<td>8</td>
<td>N=373</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>GA</td>
<td>6</td>
<td>N=156</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>NTZ</td>
<td>1</td>
<td>N=35</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

DMT = disease-modifying treatment; IFNbeta = interferon beta; GA = glatiramer acetate; im = intramuscular; NR = not reported; sc = subcutaneous; NTZ = natalizumab

### Key Points

- Overall, high risk of bias and small sample sizes rendered the evidence insufficient to address the risks of fetal exposure to DMT during pregnancy in women with MS or the risks to the mother from the drug holiday.
- Women who discontinued DMT to attempt pregnancy but did not conceive were not observed as comparison groups, nor were data gathered or grouped by time on drug holiday prior to and during pregnancy.
- IFNbeta-1a exposure did not impact spontaneous abortion rates based on the best available evidence from one high risk-of-bias study.
Detailed Synthesis

The studies were not pooled for several reasons: (1) the outcomes are relatively rare events and the studies were not appropriately powered, (2) comparison groups were heterogeneous due to lack of reporting specific types of IFNbeta, and (3) definitions of exposure differed.

Maternal risks and benefits of discontinuation were investigated by examining post-partum relapse rate. Improvement in relapse rates during pregnancy is well established. Women who discontinue DMT with the intention of becoming pregnant risk increased relapses between discontinuation and pregnancy, as well as post-partum. Given that the studied populations are those who became pregnant, none of the studies capture what happens to women who discontinue DMT but do not become pregnant. Therefore, no research has observed whether such women are at increased risk of relapse. A German study using a retrospective survey found that women stopped DMT due to intended pregnancy for a mean of 4 years, thus the period a woman with MS remains untreated is not insignificant.

Maternal Outcomes

Three small, high risk-of-bias studies reported data on the outcome of post-partum maternal relapse rate. These studies were limited by sample size and aggregation of DMTs. One study reported on natalizumab with prospective data and found no significant difference but a trend toward fewer post-partum relapses among the exposed. One study reported on glatiramer acetate and aggregated IFNbeta with mixed prospective and retrospective data and found exposed women had a lower postpartum relapse rate during the first trimester after delivery compared with the unexposed group (p<0.05). One study reported on DMT in aggregate and found women receiving any DMT at least 8 weeks during pregnancy had a significantly lower relapse rate following delivery compared with those receiving no DMT at least 3 months prior to pregnancy, with no significant differences between the exposed and unexposed groups in relapse rates during the year prior to pregnancy or during pregnancy. However, the study timing and details of the postpartum relapse rates were not clearly reported.

Fetal Outcomes

All 12 studies reported on fetal outcomes. Two studies on IFNbeta-1a, including a large case series, found that rates of spontaneous abortion did not differ significantly from population estimates. Neither study reported a significant difference in fetal death from population estimates. One small, high risk-of-bias study reported no difference in preterm delivery between the exposed and unexposed groups. Of the six studies reporting on spontaneous abortions in pregnancies exposed to aggregated IFN, one small, high risk-of-bias study found significantly higher rates of spontaneous abortion among the exposed group. This study used multivariate analysis with mixed models on an inappropriately small sample, and the population was heterogeneous, including women with non-MS diseases in both the exposed and unexposed groups. In the exposed group, the two mothers without MS discontinued DMT at 21 and 38 weeks of gestation, while the 14 mothers with MS discontinued during the first trimester.

One small, high risk-of-bias study found no overall difference in spontaneous rates between the exposed and unexposed groups. The reported rate of spontaneous abortion was statistically significantly higher among those exposed to IFNbeta 1b compared with exposure to IFNbeta 1a, glatiramer acetate, or the control group without MS (rates in these groups ranged from 3.9% to 9.1%), with as few as one or two events per group. However, the result was not significant.
compared with the MS control group (9.8%). General population estimates of spontaneous abortion are 10 to 20 percent. This outcome is subject to surveillance bias; spontaneous abortions may be unobserved, contributing to the low rates reported by many of the included studies.

No studies on IFNbeta found significant differences in fetal death rates. One small, high risk-of-bias study found an increased risk of preterm delivery with exposure to IFNbeta, based on propensity score adjustment. The study did not report the percentage of nonoverlapping subjects that were dropped from the analysis.

Six small, high risk-of-bias studies on glatiramer acetate and one small, high risk-of-bias study of natalizumab reported no significantly higher rates of spontaneous abortion, fetal death, or preterm delivery among exposed pregnancies.

A published systematic review on DMT during pregnancy among women with MS rated the evidence level and quality as good across most IFNbeta outcomes and fair for glatiramer acetate and natalizumab. However, the review based the ratings on one small study, did not report maternal outcomes, and inappropriately used an instrument for assessing systematic reviews to assess individual study risk of bias. Two other systematic reviews on pregnancy did not assess study quality or risk of bias and performed meta-analysis without regard to underpowered sample sizes or heterogeneity (such as pooling spontaneous and elective abortions). Several recent reviews have summarized FDA categorization for DMTs including oral drugs with or without current FDA approval as treatment for MS, along with early data from clinical trial registries.

KQ1c. Reasons for Discontinuation of Disease-Modifying Treatments Reported in Long-Term Observational Cohort Studies

Twenty of the 27 articles in the full reporting set reported reasons for discontinuing treatment. (Detailed evidence table is included in Appendix C.) Table 9 provides a summary of the number of articles reporting discontinuations by categories.

**Key Points**

- The broad variation in discontinuation reporting prevented useful aggregation of studies.
- All studies reported one or more adverse events and inefficacy or progression of disability as reasons to discontinue.
- Patient reasons for discontinuing DMT were not explored.

### Table 9. Studies reporting reasons for discontinuing medication

<table>
<thead>
<tr>
<th>DMT</th>
<th>Total Number of Studies</th>
<th>Adverse Event</th>
<th>Inefficacy or Progression of Disability</th>
<th>Intended Pregnancy</th>
<th>Long-Term Stable MS</th>
<th>Death</th>
<th>Protocol Violation</th>
<th>Patient Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glatiramer acetate</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Teriflunomide</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Interferon beta-1a</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Interferon beta-1b</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Interferon beta mixed</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>DMT mixed</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>
aCategory includes counts of both discontinuation based on clinician evaluation of disease progression and patient evaluation of lack of efficacy.

bProtocol violations are from studies that were RCT extensions.

**Detailed Synthesis**

As was seen with harms, the wide range of reporting methods and discontinuation categories prohibited detailed quantitative aggregation over the studies. Most articles reported numerous reasons for discontinuations. An article was categorized as reporting lack of efficacy or progression of disability based on study author definitions and use of the specific terms. In this category we included both physicians’ and patients’ perceptions of lack of efficacy, and it was often unclear who was the source of such a determination. Progression of disability was also study author defined but included articles that provided objective measures of progression such as a one-step increase in the EDSS. Protocol violations was included as a category since several studies were open label surveillance extensions of RCTs. Unfortunately, the category of most interest, the patient’s decision to discontinue, remained largely unexplored by the study authors. Minimal text was generally provided for this category, such as “by own will,”54,58 “withdrew consent,”38,60 or “voluntary withdrawal.”45,50,52 One study labeled patients stopping DMT for “no discernable reason, without discussion with their neurologists” as noncompliant.41

**KQ2. Individuals’ Values, Beliefs, and Preferences Regarding Discontinuing Disease-Modifying Treatments**

**Description of Included Studies**

The 27 included unique studies (30 total articles) represented a wide range of study aims. Designs ranged from factor analysis of questionnaires to experimental psychology lab tests to trials of shared decisionmaking interventions. Study locations were international, including the United States,94-101 Netherlands,102-104 Germany,105-116 Norway,117 a consortium of European countries,118 Canada,119,120 Italy,121,122 and Ireland.123 The included studies, organized by the KQ sub-questions and relevant section of the conceptual framework, are shown in Figure 1.

**Overarching Key Point**

Given the complexity of understanding preferences and behaviors, and the wide range of study designs used over a small literature set, all KQ2 key points should be viewed as preliminary.

**KQ2a. Patients’ and Providers’ Preferences for Discontinuation of Disease-Modifying Treatments in Patients**

Literature for this KQ sub-question is categorized as intrapersonal or interpersonal literature. Intrapersonal literature examines preferences, values, knowledge, beliefs, and other behaviors of people with MS to improve our understanding of the inputs that may enter on a personal level into deciding whether or not to discontinue DMTs. The interpersonal literature addresses the relationship between patients and providers, and the preferences each brings to the encounters and decisions.
Intrapersonal Literature

We found few studies to populate the intrapersonal portion of Figure 1 (reproduced in Figure 4). Table 10 provides information on the 14 studies that addressed intrapersonal factors in the values and preferences literature. The literature tended to examine attitudes and cognition rather than patient knowledge and how that knowledge affected decisions. Studies examined risk expectation, preferences for DMT treatment and treatment trade-offs, knowledge of cost factors, reasons for using or discontinuing DMTs, and theoretical approaches to understanding decision making and behavior processes. Study designs varied widely, including survey questionnaires, qualitative interviews, preferences, elicitation techniques, prospective cohorts, cross-sectional analysis of claims data, and one RCT. Four studies received industry funding, five did not report funding sources, one was unfunded, and four received governmental or nongovernmental funds.

Key Points

- Patients overestimated intermediate term risk of wheelchair use but underestimated the lifetime risk. This underestimation may indicate the uncertainty felt by MS patients when contemplating their personal trajectories rather than lack of knowledge (two studies).
- Patients are likely to use heuristics in risk assessments (one study).
- With training, patients can improve risk understanding and sense of informed choice (one study).
- Quantified preference studies suggest patients are willing to make risk trade-offs for benefits only to the point where the discomfort from side effects and treatment are equal to or worse than the disease symptoms (two studies).
- Increasing out-of-pocket cost reduces DMT purchases (two studies).
- Common reasons for discontinuing include side-effects, uncertainty about or perceived lack of efficacy against disease progression, administration method and frequency, and cost (five studies).
- MS patients tended to take responsibility for the decision to discontinue (three studies), while viewing their neurologist as the driver for decisions regarding choice of DMT (one study).
- Psychological models of behavior support the presence of rational processes contributing to patient decisionmaking (two studies).
<table>
<thead>
<tr>
<th>First Author, Year Country Funder</th>
<th>Study Aim</th>
<th>Population</th>
<th>Design</th>
<th>Study Author Findings</th>
</tr>
</thead>
</table>
| Janssens, 2003^1^3 | Quantify expectations of wheelchair dependency in newly diagnosed MS patients and partners | 101 MS patients diagnosed within previous 2 years, mean diagnosis time 8 months, 70% female, only 10% current DMT users; 78 partners | Survey questionnaire and neurological exams; compared perceived risk to actual at 2 years, 10 years, and lifetime; differences by clinical characteristics | • Both groups overestimate 2 and 10 year, but underestimate lifetime risk.  
• Patients with more functional limitation perceived lower seriousness for wheelchair use.  
• Concordance between patient and partner was moderate. |
| Boeije, 2004^1^2 | Examine how patients explain perception of prognostic risk | 85 MS patients included in Janssens, 2003 who were interviewed | Qualitative; one-on-one interviews in patients' homes | • Uncertainty of future progression was predominant factor in risk perception.  
• Patients discriminated over time in the perception of absolute risk.  
• Use of heuristics was clearly evident. |
| Kopke, 2014^1^6 | Evaluate efficacy of an evidence-based patient information program; increase informed choice in MS patients | 192 RRMS or CIS enrolled in RCT evaluating effectiveness of education program for informed decision making based on theory of planned behavior | RCT of complex intervention including education program presenting best available evidence regarding diagnostic testing, prognosis, and early DMT therapy. Control received stress management program by trained psychologist. Outcomes: Multidimensional measure of informed choice | • Intervention arm more likely to demonstrate good risk knowledge after 4 weeks than control (p<0.001)  
• Intervention arm more likely to achieve informed choice after 6 months compared with control (OR 0.2, 95% CI 0.1 to 0.4) |
<table>
<thead>
<tr>
<th>First Author, Year</th>
<th>Study Aim</th>
<th>Population</th>
<th>Design</th>
<th>Study Author Findings</th>
</tr>
</thead>
</table>
| Prosser, 2003       | Examine preferences for MS disability states and three DMTs. Also compare differences in utility weights between MS and healthy community | 62 RRMS patients, median age 38, 79% female; 67 healthy community respondents, similar to U.S. population but more educated and fewer children | Computer-based survey using standard gamble methods to elicit and quantify preferences for health states as input to a modeling project | • Patients discontinue DMT when side effects and discomfort from treatment are equal to or worse than disease symptoms.  
• Patients assigned higher utilities to MS-related health states and treatment states than community respondents.  
• Ratings diverged as health states worsened. For the EDSS 8 health state, there was very little overlap between patient and community respondents. |
| Johnson, 2009       | Examine willingness to accept a life-threatening adverse event in exchange for improvement | 651 MS patients drawn from consumer advocacy and industry website lists, and participants in a clinical trial; similar to national MS sample | Computer-based survey using choice-format stated preferences, also known as discrete choice experiment or conjoint analysis, to elicit and quantify preferences for health states | • Delay in years to progression was the most important factor in willingness to accept risk, followed by risk of severe adverse event; relapse rate was least important.  
• Respondents were willing to trade higher risk for greater benefit.  
• Maximum acceptable risk was relatively insensitive to SES or patient experience with MS. |
| Dor, 2010           | Examine the impact of copayments or coinsurance on reducing adherence      | 1974 MS patients with prescription or procedure codes for IFN or GA (MedStat MarketScan Commercial Claims and Encounters database of private sector health data) | Two-stage least squares model using secondary data; copayment vs. coinsurance and effects on medication possession ratio; patients with copay matched to patients with coinsurance. | • Increases in coinsurance associated with decreased adherence. 10% increase in cost sharing led to an 8.6% decline in adherence. (monthly price range $44 to $1162)  
• Copayment, which is inherently more price stable than coinsurance, did not show an effect. (monthly price range $37 to $42) |
| Gleason, 2009       | Examine the relationship between DMT abandonment and out-of-pocket expenses | 2791 MS patients with continuous enrollment and new to DMT (BlueCross BlueShield database, Midwestern and southern) | Cochran-Armitage test for trend and multivariate logistic regression; defined prescription abandonment as never taking possession of DMT | • Abandonment rate increased with increasing out-of-pocket cost. (p<0.001)  
• 5.7% abandonment rate at monthly cost of $100 or less  
• OR between 6.1 and 7.3 for expense groups greater than $200/claim (1 in 4 MS patients abandoned DMT at $200/claim) |
<table>
<thead>
<tr>
<th>First Author, Year</th>
<th>Study Aim</th>
<th>Population</th>
<th>Design</th>
<th>Study Author Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visser, 2011&lt;sup&gt;104&lt;/sup&gt;</td>
<td>Examine patient perception regarding decisions and reasons for using or not using DMT</td>
<td>1403 MS patients recruited through Dutch National MS Foundation and MS nurses (89% response rate) Includes 41% RRMS, 31% SPMS, 19% benign MS, 9% PPMS</td>
<td>Survey questionnaire, descriptive statistics; patients categorized by never used, stopped using, currently on 1&lt;sup&gt;st&lt;/sup&gt; DMT, currently on switched DMT</td>
<td>• Patients attributed the decisive role in starting (70%) or changing (66%) DMT to the neurologist, but claimed the decisive role for stopping (62%). • 1/3 of current DMT users were uncertain DMT had a beneficial effect. • Reasons for stopping: side-effects, uncertainty of benefit, administration method, aggravation of signs and symptoms, frequency of administration, and phase of disease</td>
</tr>
<tr>
<td>Bischoff, 2012&lt;sup&gt;105&lt;/sup&gt;</td>
<td>Examine reasons for discontinuation; who made and what influenced decision</td>
<td>396 MS patients discontinuing DMT in last 3 months, 75% female; 40 clinics; includes 54 SPMS on DMT</td>
<td>Telephone standardized questionnaire, descriptive statistics</td>
<td>• Mean duration of DMT was 30.5 months, ±32.1. • Positive expectations from therapy declined from 59% to 49% at discontinuation. • 75% claimed responsibility for decision; no outside influence. (Half were willing to restart.) • Reasons for stopping: not wanting to be reminded of MS, side effects, loss of efficacy, adverse events (same in all DMTs)</td>
</tr>
<tr>
<td>Grytten, 2013&lt;sup&gt;107&lt;/sup&gt;</td>
<td>Examine psychosocial factors of nonstarters and stoppers of DMT</td>
<td>424 RRMS respondents (84%), 69% female</td>
<td>Survey questionnaire, descriptive statistics and multivariate regression; patients categorized by never used, stopped using, currently on 1&lt;sup&gt;st&lt;/sup&gt; DMT, currently on switched DMT.</td>
<td>• Patient decision to stop 37%, Patient/MD decision to stop 45%, MD decision to stop 18%. • Risk of stopping increased with high education. • Psychosocial factors not significant. • Significant factors: lack of efficacy, tolerability, adverse events.</td>
</tr>
<tr>
<td>Meyniel, 2012&lt;sup&gt;108&lt;/sup&gt;</td>
<td>Assess factors leading to first treatment discontinuation</td>
<td>2314 CIS patients from 44 centers; MSBase Incident Study; 1247 started DMT (followed 2.7 years, during which 90% converted to definite MS)</td>
<td>Prospective cohort, multivariate survival analysis for factors predicting stopping</td>
<td>• 40% stopped within the observation period. • Significant factors for stopping: female, increasing EDSS, and using IFN. • Stopping varied by country • Not significant: MRI features, age, time to treatment, relapse on treatment</td>
</tr>
</tbody>
</table>
Table 10. Included studies for intrapersonal literature (continued)

<table>
<thead>
<tr>
<th>First Author, Year Country Funder</th>
<th>Study Aim</th>
<th>Population</th>
<th>Design</th>
<th>Study Author Findings</th>
</tr>
</thead>
</table>
| Daugherty, 2005<sup>97</sup> U.S. Not reported | Examine patient reported factors for discontinuing treatment | 108 MS patients prescribed DMT at a Midwest university neurology clinic. | Telephone questionnaire, descriptive statistics; a priori discontinuation reason categories | • Adverse events 52%  
• Physician-documented disease progression 40%  
• Perception of drug ineffectiveness 20%  
• Cost 4%.

| Berger, 2004<sup>96</sup> U.S. Industry | Assess extent to which Transtheoretical Model of Change predicts discontinuation | 530 (of 946, 56%) MS patients in Biogen Alliance program prescribed Avonex (INF beta 1a) 79% currently on Avonex | Survey, split into current, previous, or never users; factor analysis of readiness stages, decisional balance (pros and cons), self-efficacy; regression for predictors including cognitive impairment, injection info, demographics | • Predictors of use: average pros of DMT, average cons of DMT for decisional balance, education, and level of disability.  
• Model predicted 82% of discontinuers, 81% continuers.

| Kasper, 2012<sup>112</sup> Germany Not reported | Use theory of planned behavior to describe decision process for intention to use DMT; test evidence-based education intervention | 192 RRMS enrolled in RCT evaluating effectiveness of education program for informed decision making | Survey; validation study of measures of planned behavior factors; logistic regression using intention, attitude, subjective social norms, perceived behavioral beliefs, expectations, and values | • Planned behavior questionnaire explained 68% of variance in intention to use DMT.

CIS = clinically isolated syndrome; DMT = disease-modifying treatment; EDSS = extended disability scale score; GA = glatiramer acetate; IFN = interferon; MS = multiple sclerosis; MD = medical doctor; PML = Progressive multifocal leukoencephalopathy; QoL = quality of life; RRMS = relapsing remitting multiple sclerosis; SES = socioeconomic status
Detailed Discussion

Risk

Risk is fundamental to decisionmaking. Expectations regarding an uncertain future can figure heavily in both rational thought processes and emotional appraisals. Risk expectations have particular salience for people with MS due to the enormous prognostic uncertainty regarding the disease course. One mixed-methods study, using quantitative survey data and interviews, along with a neurological clinical examination, looked at perceptions of risk of future wheelchair use. The study did not directly examine individual choice to use or discontinue DMT, but it provides empirical literature regarding MS patient perceptions of risk. These perceptions would presumably factor into decisions regarding DMT use.

MS patients within 2 years of having been diagnosed tend to overestimate risk of becoming wheelchair-bound in the short and immediate term but underestimate the lifetime risk. The authors chose this particular risk factor because wheelchair use is one of the most major and well-recognized consequences of MS. In 101 MS patients, 2-year expected risk was 22.5 percent versus 5-10 percent actual, 10-year expected risk was 38.8 percent versus 20-25 percent actual, and lifetime risk was estimated at 54 percent versus 70-80 percent actual. People with more functional limitations were less concerned about wheelchair use than people with fewer functional limitations (p <0.01). This may reflect that MS patients with higher mobility impairment view wheelchair use as an improvement in mobility, while those with higher mobility view it as a significant loss.

Followup interviews with 85 MS patients indicated the use of heuristics—or cognitive shortcuts—to simplify complex judgments. The enormous uncertainty of the MS disease course was the predominant explanation for risk perceptions. One-third of MS patients in the quantitative study perceived their 10-year and lifetime risk of wheelchair use to be 50 percent. However, all but one patient indicating a 50 percent risk gave reasons such as “I've no idea,” or “it might happen or it might not.” This indicates that their response was based more on uncertainty about their prognosis rather than on a belief that the risk was actually 50 percent.

MS patients’ risk comprehension can be improved with education. An RCT with 192 RRMS or CIS patients receiving a 4 hour education program including MS basics, diagnostics, prognostic studies, therapeutic options, and pros and cons of DMT retained improved risk understanding after 4 weeks.

Quantified Preferences

Patients’ preferences for different health states are addressed by two studies using different methodological approaches to estimate or model the trade-offs contributing to the preferences. While some of the findings may seem evident, such as the importance of delaying disease progression, other findings are cautionary. MS patients placed more value on disabled health than did study participants without MS. Further, their preferences change over time and do so differentially for different quality of life domains.

One study used standard gamble to elicit preferences, or utilities, for health states. The standard gamble measures the risk patients are willing to take for a better outcome and the value they attach to their current health state. Patients are asked to gamble between their current certain health state and an uncertain outcome that has a probability P of full health and a corresponding probability 1-P of death. For this study, MS patients were asked to choose between a chronic treatment or a one-time treatment with some chance of perfect health and some risk of immediate...
death. Three treatment profiles were created to mimic IFNbeta-1A, IFNbeta-1b, and glatiramer acetate. The treatment descriptions did not include information on the rare but potentially serious adverse events of treatments. Respondents indicated they would discontinue DMT when the severity of side effects and discomfort from treatment are equal to or worse than disease symptoms. This creates a challenge when MS patients are early in the disease course. Such a person may not yet be experiencing significant disease symptoms, but considering treatment with negative quality of life consequences today in exchange for the possibility of delaying disease symptoms and disability. The differing preferences between MS patients and healthy people underscore that the hypothetical nature of the gamble and highlight the fact that choices made in a hypothetical situation differ from actual decisions and behaviors.

Choice-format stated-preference, also known as discrete choice experiments or conjoint analysis, was used to quantify the relative preferences MS patients assigned to various treatment attributes. This technique can estimate the maximum risk patients are willing to accept in trade for treatment benefits. This study included several major risks associated with DMTs, including death or severe disability from PML, death from liver failure, and death from leukemia. These risks were traded against different 5-year annualized relapse rates and years to disease progression. The study claimed the maximum acceptable risk estimates for the levels of benefit observed in clinical trials were higher than the observed risk but did not provide the data necessary for the comparison. The study design tried to reduce respondent burden, given the number of trade-offs to assess, and tested for result validity through mathematical rules and logic validity tests. However, as with the standard gamble, the hypothetical nature of the questions fails to inform us about actual choices. Further, respondents may have lacked the numeracy skills to fully appreciate the information provided and evaluate their choices accordingly. Heuristics may also have come into play. Although the heuristics may not have differed substantially from the actual cognitive behaviors of the decision process, heuristics itself is not consistent with the assumption of rational processes underlying this preference elicitation technique.

Cost as a Contributor to Discontinuing

Costs, especially out-of-pocket costs, have a serious impact on a patient’s decision to discontinue DMT. Two studies examined the effect of out-of-pocket costs on MS patients choosing not to fill prescriptions for DMTs. Both studies found increasing costs related to decreasing possession of DMTs. Both studies considered patient demographics (age, sex, geographic region). One study relied on zip codes to represent education and income levels. The other study matched patients by demographics plus employment status and employer industry; it also included general health status as assessed by the Charlson score and other comorbid conditions. MS patients with copayments rather than coinsurance did not exhibit a relationship between increasing cost and medication possession but still showed variation in medication possession. This finding underscores that finances are not the only factor responsible for DMT discontinuation.

Stated Reasons and Predictors for Discontinuing

Stated reasons and predictors for discontinuing DMT were examined in five studies. Four studies used surveys, while one used prospective cohort data. While all studies converged on common reasons to discontinue—side-effects, perceived lack of efficacy, aggravation of signs and symptoms/disease progression, and administration method and frequency—surveys allowing open responses, rather than a priori discontinuation categories, found more nuanced responses. For example, patients who said they did not want to
be reminded of the disease were not responding to fear of injections; patients injecting the DMT weekly were more likely to express this desire than patients injecting several times weekly or daily.\textsuperscript{105} Rather, these patients were expressing the psychological response to being confronted with the disease at the time of injection. As another example, uncertainty of DMT benefit is qualitatively different than perceived lack of benefit.\textsuperscript{104}

**Who Makes the Decision To Discontinue?**

The majority of patients took responsibility for the decision to discontinue, a finding similar in three studies that reported 62 percent,\textsuperscript{104} 75 percent,\textsuperscript{105} and even up to 82 percent if unilateral patient and patient/physician shared categories are combined.\textsuperscript{117}

**Tests of Psychological Theories**

Two studies used theories or models of behavior to explore patient preferences for DMT.\textsuperscript{96,112} The Theory of Planned Behavior, an expectancy-value model, assumes that behavior reflects past experiences and anticipated barriers. A person’s intent to perform a specific behavior depends on the individual’s favorable or unfavorable attitudes toward the behavior, perceived pressure to comply to social norms regarding the behavior, and perceived ease or difficulty of performing the behavior.\textsuperscript{125} In contrast, the Transtheoretical Model of Change assumes that behavior change is a six-step process, and people move through precontemplation, contemplation, preparation, action (new behavior for fewer than 6 months), maintenance (longer than 6 months), and termination.\textsuperscript{126} This model incorporates concepts from numerous influential theories in psychology, including motivational theory, social learning theory, self-efficacy theory, conflict decisionmaking theory, the health belief model, and the theory of reasoned action.

Kasper and colleagues\textsuperscript{112} identified domains in a questionnaire based on the Theory of Planned Behavior. These domains predicted 68 percent of the variance in the intention to use DMTs in 192 patients with RRMS enrolled in an evidence-based education intervention for informed decisionmaking, suggesting rational processes contribute to decisionmaking regarding DMTs. The assessed domains included attitudes, social norms, and beliefs regarding the perceived ease or ability of the patient to perform expected behaviors. The questionnaire also incorporated emotional appraisals of DMTs, through questions such as “the risk I would be taking by putting off immunotherapy for too long frightens me.” However, the study examined intention to use DMT, not the actual behavior.

Rational processes also contributed to decisionmaking when modeled using the Transtheoretical Model of Change.\textsuperscript{96} Along with level of education and disability, average pros and average cons of DMT use, specifically IFNbeta-1a in this case, predicted 81 percent of discontinuation. The study authors determined an equation for calculating an individual’s likelihood of discontinuing treatment. Very simply, as patients lose confidence in the DMT’s ability to prevent or slow disease progression, they move closer to discontinuing. The study was conducted by one firm to develop a company-based program to promote treatment persistence, so generalizability of the findings is uncertain.
Interpersonal Literature

Much less literature populates the interpersonal portion of Figure 1 (reproduced in Figure 5). Table 11 provides information on the three studies that addressed interpersonal factors in the values and preferences literature. Interpersonal concerns include the knowledge, values, beliefs, and preferences that both the patient and physician bring to a decisionmaking encounter, and also the extent to which this information is shared between the two. Communication issues also are important at the interpersonal level. Study designs included survey questionnaires and qualitative interviews. One study received industry funding, one was unfunded, and one received nongovernmental funds.

Key Points

- MS patients and their physicians can differ significantly in their perceptions of the relative importance of health states and risks (two studies).
- Physicians and patients must communicate in order to clarify differences in perceptions and preferences (one study).
<table>
<thead>
<tr>
<th>First Author, Year, Country, Funder</th>
<th>Study Aim</th>
<th>Population</th>
<th>Design</th>
<th>Study Author Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heessen, 2010&lt;sup&gt;10&lt;/sup&gt; Germany Not funded</td>
<td>Assess risk tolerance of MS patients and treating physicians for natalizumab adverse events</td>
<td>69 MS patients from a university clinic (average EDSS = 4), 192 affiliated neurologists</td>
<td>Survey to investigate prerequisites for shared decisionmaking (part of an intervention study of shared decisionmaking). A 3-page information leaflet on natalizumab was provided to both groups. Test and correlation statistics</td>
<td>• 49% of physicians would stop natalizumab for a PML risk of 2 per 10,000; only 17% of patients would do so (p&lt;0.001). • Both groups overestimated natalizumab treatment effects. • Similar to risk perceptions, patients were more strongly in favor of continuing than physicians, who were more ambivalent</td>
</tr>
<tr>
<td>Kremenchutzky, 2013&lt;sup&gt;19&lt;/sup&gt; Canada Biogen</td>
<td>Compare neurologist and MS patient perceptions of MS related health status</td>
<td>99 MS patient and neurologist pairs. 6 clinics in Ontario and Alberta; excluded patients receiving natalizumab or IFNbeta marketed by Biogen</td>
<td>Survey questionnaire using standard gamble methods to elicit and quantify preferences for health states to assess communication</td>
<td>• Significant differences between patient and provider perceptions of relapse frequency and QoL; patients rate as worse. • Little concordance on identified important health domains; providers identify physical functioning while patients emphasize mental health</td>
</tr>
<tr>
<td>Thorne, 2004&lt;sup&gt;20&lt;/sup&gt; Canada Nongovernmental</td>
<td>Describe health care communication issues</td>
<td>12 long-term MS patients, 7 unemployed due to disability; prior to illness, 6 had been professionals, 3 health care professions</td>
<td>In-depth, loosely structured interviews, recorded, transcribed, and analyzed using NVivo software.</td>
<td>• Provides table of helpful and unhelpful communication specifics</td>
</tr>
</tbody>
</table>

EDSS = extended disability scale score; MS = multiple sclerosis; PML = Progressive multifocal leukoencephalopathy; QoL = quality of life
Detailed Discussion

Differences Between Patient and Physician Perceptions

MS patients and their physicians may differ in their perceptions of the relative importance of health states and risks. Two studies examined these differences, one using a survey, the other using the standard gamble to quantify preferences. Standard gamble was described briefly in the section on intrapersonal literature. Both studies found differences between patient and physician relative assessments of risks and health states. The lack of concordance carried into preferences to discontinue DMT for given risk levels for serious adverse events, in this case, discontinuing natalizumab for given risk levels for PML. Patients had a more negative perception of MS than physicians and were willing to tolerate greater risk of PML. Physicians tended to value health states more highly than patients, and emphasized physical health states, while patients gave greater relative weight to mental health. With only six neurologists for the 99 patients in this study, generalizability is low. Further, the MS patients were recently diagnosed and thus had limited experience with their condition. Possibly, the discord between physician and MS patients would have lessened with a more experienced group of long-term MS patients. However, these studies highlight the need for good communication between physician and patient for successful shared decisionmaking.

One study examined communication issues from the perspective of the patient using qualitative interviews of 12 long-term MS patients. While the study was not strictly related to DMT use and discontinuation, we include it in this section as a resource for communication based on empirical research. The table of communication tips from that study is recreated in Table 12.

Table 12. Helpful and unhelpful communications in MS

<table>
<thead>
<tr>
<th>Coping Focus</th>
<th>Helpful Communication</th>
<th>Unhelpful Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing Fear</td>
<td>Timely, relevant, accurate information</td>
<td>Withholding information</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Using statistics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sugar coating</td>
</tr>
<tr>
<td></td>
<td>Validating patient experience</td>
<td>Dismissing patient claims</td>
</tr>
<tr>
<td></td>
<td>Planned appointments to discuss test results, ongoing management, and follow up</td>
<td>Waiting for referrals and appointments</td>
</tr>
<tr>
<td>Taking Charge</td>
<td>Assistance from health care professionals</td>
<td>Feeling alone and isolated in managing MS</td>
</tr>
<tr>
<td></td>
<td>Providing as much information as possible – more is better than less</td>
<td>Inaccurate or outdated information</td>
</tr>
<tr>
<td></td>
<td>Convenient access to medical professionals</td>
<td>Difficulty accessing medical services</td>
</tr>
<tr>
<td></td>
<td>Acknowledging the limits of medical science</td>
<td>Belief that medical science has all the answers</td>
</tr>
<tr>
<td>Crafting a Life</td>
<td>MS is only one aspect of their life</td>
<td>MS is their life</td>
</tr>
<tr>
<td></td>
<td>Validating symptoms</td>
<td>Minimizing symptoms</td>
</tr>
<tr>
<td></td>
<td>Ongoing interest</td>
<td>Giving up</td>
</tr>
<tr>
<td></td>
<td>Willingness to learn and explore alternative therapeutic options</td>
<td>Inflexibility in thinking and researching</td>
</tr>
<tr>
<td></td>
<td>Respecting the patient as a competent and knowledgeable partner</td>
<td>Condescension</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Platitudes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reprimanding</td>
</tr>
</tbody>
</table>

MS = multiple sclerosis

Table adapted from Thorne, 2004.
KQ2b. Patients’ and Providers’ Preferences for Participation in Shared Decisionmaking To Discontinue Disease-Modifying Treatments

Literature for this KQ subquestion relates to shared decisionmaking for patients and providers. All but one of the identified studies populate the center box in the shared decisionmaking portion of Figure 1 (reproduced in Figure 6). Table 13 provides information on the five studies addressing shared decisionmaking from the patient side, four addressing the physician side, and a test of a decision aid to improve shared decisionmaking. Study designs included survey questionnaires, experimental psychology, third party observation of physician shared decisionmaking skills, and a randomized controlled trial of the decision aid. Six studies received governmental or nongovernmental funding, four did not report funding, and one was unfunded.

Key Points

- MS patients may bring different information-seeking orientations to shared decisionmaking (one study).
- Mildly cognitively impaired MS patients show a significantly reduced capacity to understand treatment disclosures, but understanding may be brought back to the level of healthy controls through repetition and recognition cuing (one study).
- The large majority of people with MS prefer a collaborative or active role in treatment decisions (three non-U.S. studies).
- Physicians cannot reliably predict patient preferences for an active participation role and may inadvertently pull patients away from their preferred treatment (two studies).
- Both patient and third party observers rated physicians as showing limited skill at involving patients in shared decisionmaking (one study).
- Providing balanced, evidence-based information alone is not sufficient to alter decisionmaking processes to help patients achieve their preferred participation role (one study).
<table>
<thead>
<tr>
<th>First Author, Year Funder</th>
<th>Study Aim</th>
<th>Population</th>
<th>Design</th>
<th>Study Author Findings</th>
</tr>
</thead>
</table>
| Baker, 1995<sup>1st</sup>  
U.S.  
Not reported | Test if general orientation and length of MS affects interest in and desire for (1) more information about MS, (2) general or specific information | 1. 160 women with MS  
2. 95 women with MS | Survey questionnaire to identify monitors vs. information blusters, comparative tests of  
1. Interest in and amount wanted for 29 MS topics  
2. Relevance of information | • Monitors want both general and specific information from the beginning.  
• Blunters want information for some topics only after having MS for period of time. |
| Basso, 2010<sup>8th</sup>  
U.S.  
Governmental and nongovernmental | Assess whether aspects of neurocognition correspond with understanding treatment disclosures, and if disclosure understanding can be enhanced | Experimental: 36 people with MS; 24 unimpaired, 12 cognitively compromised  
Control: 16 adult community participants without MS | Experimental psychology lab  
Measures of new learning, executive function, attention  
Understanding of treatment disclosures scale. Significance tests and correlations | • Lower new learning and executive function correlated with poor understanding of treatment disclosures; understanding about 60% of control group mean.  
• Repetition and cuing improved understanding of cognitively compromised back to the level of control adults  
• MS patients without cognitive impairments understood information as well as the control group. |
| Giordano, 2008<sup>14th</sup>  
Italy  
Nongovernmental | Assess preferences of Italians with MS regarding participating in treatment decisions | 129 people with stable MS, 71% female, 69% RRMS, average EDSS 3.5. | Survey using Control Preference Scale. Sex, age, disease duration, EDSS score, education level, whether on DMT, length of followup time at clinic. | • 61% preferred a collaborative role, 33% passive role, 6% active.  
• Sex, age, disease duration, EDSS score, and whether on DMT were not associated with preferred role.  
• Those with ≥5 years followup time at the clinic were more likely to prefer passive. |
<table>
<thead>
<tr>
<th>First Author, Year Funder</th>
<th>Study Aim</th>
<th>Population</th>
<th>Design</th>
<th>Study Author Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamann, 2007 (^{107}) Germany Governmental</td>
<td>Assess if decisionmaking participation preferences vary by chronicity of disease</td>
<td>1393 German patients with hypertension (164), depression (230), breast cancer (178), Schizophrenia (120), RRMS (105), Minor traumas (596)</td>
<td>Pooled preferences from SDM trials. Examined age, gender, education, diagnosis, and autonomy preference index scale Descriptive and test statistics</td>
<td>• MS patients significantly more likely to have higher preferences for participation than other patient groups (p &lt;0.001)</td>
</tr>
<tr>
<td>Heesen, 2007 (^{108,109,111}) Germany Governmental</td>
<td>Assess decision role preferences and MS risk knowledge</td>
<td>113 RRMS and 100 PPMS patients randomly selected from MS database</td>
<td>Survey questionnaire (performed during decision aid development) Descriptive and test statistics</td>
<td>• 79% (132/168) prefer an active role in decisionmaking. • Patients on INF had higher knowledge (risk calculation ability) than those not on INF. • People who preferred an active role had higher knowledge</td>
</tr>
<tr>
<td>Hamann, 2010 (^{106}) Germany Not reported</td>
<td>Examine how accurately neurologists and psychiatrists can predict patient participation preferences</td>
<td>203 patients; 102 MS patients, 66% female, mean MS duration 10 years; 101 schizophrenia, 51 providers</td>
<td>Survey questionnaire Autonomy preference index scale. Descriptive and test statistics</td>
<td>• Physicians tended to over predict patient preference for participation • Physicians correctly predicted 74% of preferences, but agreement between physician and patient was overall poor. • Participation preferences depended on the physicians’ and patients’ expertise with MS</td>
</tr>
<tr>
<td>Pietrolongo, 2013 (^{122}) Italy Nongovernmental</td>
<td>Assess physician shared decisionmaking from third party observer and patient perspectives</td>
<td>88 MS patients, 10 physicians</td>
<td>Audio recordings of first consultations rated using Observing Patient Involvement in Shared Decision Making tool</td>
<td>• Physicians were rated as showing limited skills at patient involvement by both patients and third party observers; patients rated physicians higher than the third party observer.</td>
</tr>
<tr>
<td>Mendel, 2011 (^{178}) Germany Not funded</td>
<td>Examine if physicians influence patients away from their preferred treatment option.</td>
<td>102 MS patients, 66% female, mean MS duration 10 years</td>
<td>Fictitious clinician recommendation contrary to stated preference Stated choice, satisfaction with choice</td>
<td>• 26% of MS patients followed the physician’s advice, but were less satisfied with their choice.</td>
</tr>
<tr>
<td>First Author, Year Funder</td>
<td>Study Aim</td>
<td>Population</td>
<td>Design</td>
<td>Study Author Findings</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------</td>
<td>------------</td>
<td>--------</td>
<td>-----------------------</td>
</tr>
</tbody>
</table>
| Lonergan, 2009\(^{123}\) | Assess utilization of DMT in patients with SPMS and PPMS, and examine approaches by physicians to counsel discontinuation in SPMS | 336 Irish patients with MS living in urban southeast Dublin, and rural counties Wexford and Donegal. 26 neurologists in Europe, North America, and New Zealand | Survey questionnaires to physicians. Patients recruited through clinical services and MS Society, medical chart review for utilization. | • 27% of patients with EDSS >6.5 were using DMT and were PPMS or SPMS, 99% of which were in rural locations.  
• 15 physicians made an effort to stop treatment; 2 never stopped prescribing, and 11 generally continued DMT.  
• Despite stated preference to stop, most did not insist to avoid affecting the relationship. |
| Decision Aids | Kasper, 2008\(^{13,14}\) | Evaluate effects of a patient decision aid to improve achieving a preferred role in decisionmaking and DMT use | 297 MS patients considering or reconsidering DMT. 14% CIS, 53% RRMS, 19% SPMS, 11% PPMS. Recruited from community, consumer advocate group, MS clinic. Experimental: 150, Control: 147 | RCT of evidence-based decision aid. Control received standard treatment information. Outcomes: Achieved participation role preference; treatment choice | • No differences between groups on role preference or treatment choice.  
• 50% of patients achieved preferred role. Patients tended to use shared decisionmaking as an actual role.  
• 18% of patients reconsidering treatment chose to interrupt treatment.  
• Experimental group was more critical of DMT use than control group. |

CIS = clinically isolated syndrome; DMT = disease-modifying treatment; EDSS = extended disability scale score; MS = multiple sclerosis; PPMS = primary progressive multiple sclerosis; RCT = randomized controlled trial; RRMS = relapsing remitting multiple sclerosis; SDM = shared decisionmaking; SPMS = secondary progressive multiple sclerosis
Detailed Discussion

Patients in Shared Decisionmaking

Two studies examined information-seeking and information-processing, necessary steps in shared decisionmaking. The type and timing of information sought by women with MS differed depending on the individual orientation toward information—whether the person was a monitor or blunter of information. Monitors prefer a high information input before a stressful event and suffer less psycho-physiological arousal when they have information while blunters prefer less information and suffer more arousal when they have a high information input. A person’s tendency toward monitoring or blunting differed according to time from diagnosis: monitors wanted information early while blunters’ preference for information that was specific increased with the time they had MS. This was a small study of women with MS, so the applicability of the findings is highly constrained. However, it does suggest that patients bring different orientations to the information search aspect of shared decisionmaking.

Cognitively compromised MS patients had impaired understanding of treatment disclosures in one study using experimental psychology lab methods. Cognitively unimpaired MS patients understood treatment disclosures as well as people without MS. Among those with cognitive impairment, probing for understanding, repeating information, and recognition cuing brought understanding performance back to the level of the healthy control group. The cognitively impaired group was relatively small and the impairment was generally mild rather than severe, thus limiting applicability. Yet, the mildly compromised MS patient may be less likely to be recognized as having reduced decisional capacity for informed consent and shared decisionmaking.

Three studies from Italy and Germany examined patient preferences for participating in shared decisionmaking. Overall, the large majority of people with MS preferred an active participatory role in treatment decisions (68% to 79%). People with MS were significantly more likely than people with hypertension, major depression, breast cancer, or schizophrenia to prefer an active role. Those who preferred an active role also tended to have higher knowledge regarding risk. The applicability of these studies to MS patients in the United States is difficult to judge without U.S. studies for comparison.

Physicians in Shared Decisionmaking

Two studies examined physician ability to predict MS patient preferences to participate in treatment decisions. Both studies used the same MS patient population in Germany. Physicians tended to overestimate patient preference for participation overall, and individual predictions correlated poorly to individual patient preferences. In a test using fictitious clinician’s recommendations, 26 percent of MS patients followed the fictitious recommendation and chose the treatment option that went against their initial preferences. A third study found both third party observers and patients rated physicians as showing limited skill at involving patients in shared decisionmaking.

One study surveyed an international sample of physicians regarding their experience with discontinuing DMT when it is no longer effective. While primarily influenced by absence of relapses and evidence of disease progression, EDSS scores in the 6.0 to 8.0 range were cited as indicators of likely lack of efficacy, although reimbursement policies in several countries were also influential. Most physicians noted the challenge of negotiating discontinuing a DMT that is
well-tolerated even if it appeared to the physician to lack efficacy. Physicians tended to defer to patient preferences for discontinuing, noting the importance of a good patient-physician relationship. Several mentioned that patients expressed insecurity with discontinuing.

**Decision Aids for Shared Decisionmaking**

Simply providing patients with balanced information did not in itself alter the decisionmaking process, although providing evidence-based information to MS patients improved assessment of information, with the experimental group more critical of DMTs than the control group.\(^{113,114}\) In one RCT of a decision aid to improve patient achievement of their preferred participation role, MS patients had a high preference rate for active or autonomous decisionmaking roles, 79 percent at baseline and 81 percent at followup in the experimental group. Actual participation roles, however, tended toward shared decisionmaking styles. In an applied context, 27 percent of the patients with stated preferences for autonomy participated at a shared level. No major differences were seen in achieving preferred roles between the intervention and control groups at followup. Across the groups, half of the MS patients reported an actual participation role other than their preferred role.

Patients and assessors were masked as to whether the patient received the decision aid or standard information packet. Patients were asked to not directly discuss the details of the information they received with their physicians. It is not possible to determine if factors related to the masking itself may have resulted in patients feeling constrained in communicating with their physician.
Discussion

Overview

Effective health care relies on the three legs of physician clinical experience, patients’ knowledge of their specific health situations and preferences, and an evidence base. Together, these three components provide the input for medical decisions. In the absence of a clear, unambiguous path to follow, patients are best served by shared decisionmaking, which requires clinicians to provide the best available information against which patients can weigh their preferences and risk tolerance.

The decisions around discontinuation of DMT treatment are extremely personal and individual. It is hard to envision ever having enough information to cover all contingencies. Providers and MS patients who have followed a prolonged DMT treatment plan have little information to guide decisions regarding discontinuing DMT. Thus, personal preferences about risks take on more weight.

No literature directly compared continuing versus discontinuing DMT in comparable populations. Only sparse information was available to address one part of the decisionmaking picture faced by providers and patients, long-term benefits and harms. Of 27 unique studies identified, only 16 provided sufficient information to adequately assess outcomes and risk of bias. As summarized in Table 14, low-strength evidence showed an increased all-cause mortality for patients who started interferon beta 1b 2 years earlier than the comparators, but no differences between treated and comparator groups in time to progression to secondary progressive MS (SPMS) (as measured by sustained Expanded Disability Status Scale (EDSS) greater than 6). Similarly, overall long-term harms were found to be no different than short-term harms. Low-strength evidence implies low confidence in the findings and the expectation that future research could change the findings. Evidence is insufficient to assess long-term benefits and harms for all other patient populations, type of DMT, or outcome.

Table 14. Summary of KQ1 findings with sufficient evidence

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>DMTs Used in Long-Term Studies Assessing Discontinuing or Continuing DMTs</th>
<th>Number of Studies Number of Participants</th>
<th>Findings</th>
<th>Strength of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-Cause Survival</td>
<td>Interferon beta 1b</td>
<td>1 study&lt;sup&gt;35&lt;/sup&gt; n=366 RRMS</td>
<td>All-cause mortality: 250 mg arm vs. placebo – HR 0.532 (98% CI, 0.31 to 0.90) 50 mg arm vs. placebo – HR 0.54 (95% CI, 0.32 to 0.92) favors treatment</td>
<td>Low (moderate risk of bias, unknown consistency)</td>
</tr>
<tr>
<td>Time to Progression to SPMS</td>
<td>Mixed Interferon</td>
<td>1 study&lt;sup&gt;37&lt;/sup&gt; n=2656 RRMS</td>
<td>No difference from contemporary or historical control</td>
<td>Low (moderate risk of bias, unknown consistency)</td>
</tr>
<tr>
<td>Overall Harms</td>
<td>Interferon, glatiramer acetate, teriflunomide</td>
<td>3 studies&lt;sup&gt;35,36,38,39&lt;/sup&gt; n=746 RRMS interferon beta 1b, 16y; 46 RRMS glatiramer acetate, 22 y; 131 RRMS, 16 SPMS, teriflunomide, 8.5y</td>
<td>Long-term harms not different than short term (qualitative finding)</td>
<td>Low (high risk of bias, consistent, indeterminate precision)</td>
</tr>
</tbody>
</table>

DMT = disease-modifying treatment; HR = hazard ratio; KQ = Key Question; RRMS = relapsing remitting multiple sclerosis; SPMS = secondary-progressive multiple sclerosis; y = year
The current literature did not examine whether long-term benefits for DMTs remain after a patient converts to SPMS. Such a finding would refute the practice of discontinuing current DMTs once a patient converts. However, given that mitoxantrone (with a maximum lifetime dose limit) is the only FDA-approved DMT for SPMS, there is little to suggest a long-term study past conversion to SPMS would be constructive. This assumption is supported by a recent systematic review and meta-analysis of 3082 SPMS patients, which found interferon beta did not reduce disability progression, and while there was a slight reduction in number of patients who had relapses during the 3 years, more treated than placebo patients dropped out due to adverse events.\(^{127}\)

For the special cases of natalizumab and planned pregnancy discontinuations, evidence was insufficient to answer whether discontinuation is problem-free. Evidence from eight studies was insufficient to address whether rebound due to discontinuing natalizumab exists. Studies lacked quantitative definitions of rebound, were high risk of bias, small sample sizes, and potential plausible confounders for post-natalizumab annualized relapse rates. Similarly, evidence from 12 studies was insufficient to address risks of fetal exposure to DMT during pregnancy in women with MS or the risks to the mother for the drug holiday due to study high risk of bias and small sample sizes. Further, women who discontinued DMT to attempt pregnancy who didn’t conceive were not observed as comparison groups, nor were data gathered or grouped by time on drug holiday prior to and during pregnancy.

Harms from long-term DMT use, as is true for many treatments and medical conditions, is poorly reported in the literature. The low-strength evidence showing long-term harms to be generally similar to short-term may very well be upset by improved tracking and reporting. For example, a November, 2014 FDA Drug Safety Communication reported the first confirmed fatal case of PML for an MS patient using dimethyl fumarate (Tecfidera).\(^{128}\) The patient had used dimethyl fumarate for 4 years.

In the absence of evidence, providers and patients are left with little to inform their preferences and guide their decisions regarding if or when to discontinue treatments. The majority of included studies reported reasons for patients discontinuing treatments, but the information provided was without detail. Adverse events and inefficacy or progression of disability were two expected categories. Other possible reasons for discontinuation, such as a patient’s desire to try alternative medicine approaches, perceived risk of long-term use, or financial concerns such as out of pocket costs or loss of insurance, are not noted. The “patient decision” category for discontinuing was consistently unexplored.

Key Question (KQ) 2 aimed to delve into what is known about patient and provider preferences. While the literature was sparse, with only 28 studies available to populate the conceptual map provided in Figure 1, each of the three major conceptual areas was at least partially represented. No study directly asked why people are reluctant to discontinue when treatment seems no longer effective, but taken as a whole the literature set provides some insight.

Overall, one can weave together the general themes found in the KQ2 literature. Admittedly, physicians cannot reliably predict patient preferences for shared decisionmaking and often perceive the relative importance of health status or acceptable risks differently. However, when it comes to the decision to discontinue DMT, the patient drives the decision, and this preference and role are generally unchallenged by the physician. In some DMT discontinuations, the balance of shared decisionmaking may shift to discordance between the physician and patient, with the physician deferring to patient preferences for continuing treatment or not. The quantified preferences work by Prosser and colleagues\(^{101}\) illustrates a paradox, where patients are
less likely to prefer DMT during the early course of the disease, when disease symptoms are lower than the side effects of the DMT, and more likely to use it at later stages of the disease when the side effects are less than disease symptoms. This behavior is counter to the hypothesis under which DMTs are assumed to work, which is by reducing relapses early in the disease course to prevent or delay disease progression. Without more solid evidence for the long-term net benefits, or the thresholds at which treatment is no longer effective in preventing disease progression, the decision to discontinue treatment remains preference sensitive.

The preferences literature underscores the complexity of the topic and the processes underlying decisionmaking. Both rational and nonrational (such as heuristics) processes came into play, and neither had primacy over the other. Cost was a factor in both self-report and through observation of purchasing behavior. Cognitive deficits impairing decisional capacity may be overcome with adequate cuing. Information is a necessary component of decisionmaking, yet nonrational factors can influence what information is sought at what time.

Preferences, values, and beliefs are highly variable, may change over time, and are linked to the nature of the patients’ relationships with their doctors. There may well be differences based on age, sex, race, class, and other factors. A patient’s preference position between “treat my MS at any cost/comfort from knowledge of receiving treatment” and “need strong evidence that the medication will help and be worth the cost/side effects” may change over time and as the disease changes.

Changing perceptions regarding health states was common across different parts of the intrapersonal literature. Risk perceptions and quantified preferences (which are risk-based as well) both suggested that people with longer MS experience assigned higher values to, or viewed as less serious, disabled states. This is a finding consistent with other research into how people value different health states. Many people overestimate their aversion to hypothetical states of disability and hence eliminate treatment options that might lead to such disability, especially if it could be long-term. The hypothetical disutilities for these states are consistently higher than for those actually experiencing the state.

Issues

Several challenges impede the gathering of evidence to inform decisions to discontinue DMTs. First, the potential differential effectiveness of DMTs for different patient subpopulations is unclear, due both to lack of studies examining the questions as well as the use of unsatisfactory study designs. Whether DMTs for clinically isolated syndrome (CIS) patients is effective remains an open question. DMTs may offer little benefit in exchange for side effects and potential harms for patients with a benign MS course. Conversely, who is at risk of worsened disease activity (such as a rebound effect or overshoot) when DMTs are discontinued, possibly prematurely? We cannot currently predict early or benign disease courses.

Second, similar to determining which CIS patients will convert to MS, or which MS patients will have a benign disease course without use of DMTs, the transition from relapsing-remitting MS (RRMS) to SPMS is difficult to ascertain and therefore poses challenges in the decision to discontinue treatment. Clear biomarkers do not exist, and neither do distinct boundaries for the transition. Currently, EDSS changes, or a score of 6 or 7, and clinical judgment are generally used. Furthermore, some patients with RRMS never transition to a clear secondary progressive phase. Since relapses tend to decrease in frequency with advancing age (they are rare after the sixth decade of life, and very rare after the seventh decade), the problem arises in determining whether a patient's lack of relapses is due to ongoing DMT or to the natural history of the
disease. For example, consider a 75-year old patient who developed RRMS at age 30, has been taking DMT since 1994, has had no relapses or new MRI lesions since 1996, and has shown no evidence of secondary progression (stable EDSS). Is the lack of relapses due to ongoing DMT use, or has this patient's MS reached the stage where the risk of relapse is passed, yet there is no ongoing neurologic deterioration beyond what would be expected in normal aging (sometimes referred to as "burned out" MS)? Is it safe to discontinue DMT in such patients? Adequate data to answer this question are not yet available.

This observation leads to the third major challenge, measuring disability. The EDSS is the most commonly used scale in research, in part because it is the longest standing. Because the EDSS is largely driven by mobility assessment, available research is generally silent on potential benefits of DMT other than ambulation, such as upper limb function and cognitive impairments. Other validated measures of health status in MS that incorporate more function domains include the MSQOL-54, the Functional Assessment of Multiple Sclerosis (FAMS), and the Multiple sclerosis Quality-of-Life Inventory (MSQLI). Since the EDSS is often used as a validation standard, cross-walks between the EDSS and other scales are already somewhat established to assist with aggregating findings across studies for mobility disability and possibly inform transitions to SPMS. However, as seen in KQ2, given that people with MS can value health domains differently than physicians (or perhaps researchers), the broader range of disability assessment should be pursued regardless of any potential limitations comparing results with studies that used the EDSS exclusively.

Without adequate measures of quality of life, balancing the benefits of treatments against harms becomes challenging, especially across different drug regimens. DMTs are not benign with regard to side effects and risk profiles. The degree to which quality of life benefits of treatment are offset by quality of life decreases due to side effects and risk profiles is important. Such research is often done within the context of cost-benefit analysis, a methodology with its own set of strengths and limitations.

Much remains to be done to understand patient preferences. Emerging but useful information was available to explore KQ2, but no study directly asked the question about preference for discontinuing treatment or explored why patients may be unwilling to discontinue even when treatment no longer appears effective. Lonergan and colleagues approached the question tangentially, asking physicians about how they counsel patients when considering discontinuation. Providers who are involved with such counseling sessions would also benefit from research that separates understanding of preferences, which may be clear to the patient, and the mixed feelings such preferences may generate, ranging from fear or grief related to “giving up” on the disease to relief for no longer carrying the burden of DMTs.

Newly approved drugs, such as fingolimod, and drugs in the development pipeline are emphasizing oral administration, which may improve medication uptake and adherence to treatment programs. Self-injection can be a deterrent to patients with MS starting injectable DMTs and “shot-fatigue” is a significant factor for adherence. Oral medications will certainly have implications for preferences for continuing and discontinuing DMTs.

Future Research

Since only three areas of evidence for KQ1 were sufficient to provide answers with low strength of evidence, essentially all questions related to KQ1 would benefit from further study. The utility of studies for estimating long-term treatment effectiveness in MS can be improved by using prospective, population-based designs with appropriate comparators and standardized data
collection methods. Study cohorts must be better characterized with respect to demographic and clinical characteristics, as well as other factors that may influence outcome such as socioeconomic status, access to care, health behaviors, and comorbidities. Near-complete patient retention with regularly scheduled patient visits is also necessary. The ability to account for treatment effects would improve with better models to predict disability outcomes in MS, including disentangling the young versus old from the new versus long-term disease presence, since the two overlap. Techniques to adjust for selection bias such as regression analysis or propensity scores are more easily accomplished with rich datasets. With regard to the question of discontinuing for pregnancy, appropriate comparison groups need to include women who discontinued DMT to attempt pregnancy but didn’t conceive. Since the pharmaceutical industry would not benefit from strong comparator studies focusing on treatment discontinuation, other funding sources will need to be identified.

Some current efforts to improve longer-term research do exist; for example, the prospective, 5 year OPT-Up study. While more geared toward initial treatment and switching choices, understanding discontinuation within that context is one of the study goals. Another prospective 10-year observational study based on the United Kingdom’s MS risk sharing scheme is evaluating the effectiveness of the first DMTs, interferon and glatiramer acetate. After NICE recommended against DMTs in 2002, a pricing scheme was negotiated with participating pharmaceutical companies whereby the drug prices would be reduced if patient outcomes were lower than expected, thus the United Kingdom National Health Service and the pharmaceutical companies shared the financial risk for cost-effective treatment. The initial 2-year results published in 2009 found patient outcomes were worse than predicted. However, results were controversial; an independent group that reviewed the data concluded that both the control dataset and analysis model selected when setting up the risk-sharing scheme, had intrinsic flaws. Four-, 6-, and 8-year data have been collected and are being analyzed using an updated modeling methodology. This research initiative should help inform the long-term benefits of these injectable treatments and may suggest improvements to current MS registries or methods, making analysis of such registries more fruitful.

KQ2 covered a broad array of relevant topics, and investigator-driven research remains a likely source for innovative and interesting approaches to continued exploration. The AutoMS project, an international consortium of six European locations and Australia, was formed in 2010 to explore MS patient preferences for shared decisionmaking. Confirming the generalizability of their findings to the United States would be beneficial. Also useful would be well-designed qualitative and survey research, perhaps using mixed-methods approaches, aimed at exploring why and under what circumstances a patient might seek to terminate treatment, and why people may be reluctant to discontinue when treatment appears no longer effective.

Attention to areas such as implementation science and quality improvement, which combined are often referred to as healthcare improvement science, should be included in the future of MS research and the improvement of MS treatment decision and outcomes. Such areas may contribute systems-level factors to DMT selection and adherence, and to the successful implementation of shared decisionmaking.

**Limitations**

Literature on preferences is not indexed to permit easy identification of relevant articles. Search strategies to capture the diffuse literature used natural language as keywords. While we tested multiple terms before settling on the final algorithm, relevant articles were likely missed,
and thus the included literature set must be viewed as comprehensive but not exhaustive. Likewise, setting the review scope to exclude adherence literature – as adherence by definition connotes a decision to continue DMT use – may have precluded some relevant literature examining lack of adherence as a de facto decision to discontinue use.
References


55


<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARR</td>
<td>Annual relapse rate</td>
</tr>
<tr>
<td>BREMS</td>
<td>Bayesian risk estimate for MS</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>Cochrane Central Register of Controlled Trials</td>
</tr>
<tr>
<td>CER</td>
<td>Comparative effectiveness review</td>
</tr>
<tr>
<td>CIS</td>
<td>Clinically isolated syndrome</td>
</tr>
<tr>
<td>DMT</td>
<td>Disease-modifying treatment</td>
</tr>
<tr>
<td>EDSS</td>
<td>Expanded Disability Status Scale</td>
</tr>
<tr>
<td>FAMS</td>
<td>Functional Assessment of Multiple Sclerosis</td>
</tr>
<tr>
<td>GEL</td>
<td>Gadolinium-enhancing lesion</td>
</tr>
<tr>
<td>HSRProj</td>
<td>Health Services Research Projects in Progress</td>
</tr>
<tr>
<td>ICTRP</td>
<td>International Controlled Trials Registry Platform</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>KQ</td>
<td>Key Question</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MS</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>MSQLI</td>
<td>Multiple Sclerosis Quality-of-Life Inventory</td>
</tr>
<tr>
<td>MSQOL</td>
<td>Multiple sclerosis quality of life</td>
</tr>
<tr>
<td>PICOTS</td>
<td>Population, intervention, comparator, outcomes, timing, setting</td>
</tr>
<tr>
<td>PML</td>
<td>Progressive multifocal leukoencephalopathy</td>
</tr>
<tr>
<td>PPMS</td>
<td>Primary progressive MS</td>
</tr>
<tr>
<td>PRMS</td>
<td>Primary relapsing MS</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>RRMS</td>
<td>Relapsing-remitting MS</td>
</tr>
<tr>
<td>SIP</td>
<td>Scientific information packet</td>
</tr>
<tr>
<td>SPMS</td>
<td>Secondary progressive MS</td>
</tr>
<tr>
<td>TEP</td>
<td>Technical Expert Panel</td>
</tr>
<tr>
<td>TOO</td>
<td>Task Order Officer</td>
</tr>
</tbody>
</table>
Appendix A. Search Algorithms

For KQ1, we searched Medline via Ovid, Cochrane Libraries, and Scopus, modifying the Medline searches for the other databases.

KQ1. MS/Drug Holiday:
Database: Ovid MEDLINE
Search Strategy:

```
1  exp multiple sclerosis/dt, th, im
2  drug holiday$.mp.
3  discontinu$.mp.
4  halt$.mp.
5  cessat$.mp.
6  interrupt$.mp.
7  stop$.mp.
8  2 or 3 or 4 or 5 or 6 or 7
9  1 and 8
```

KQ1. MS/Immunomodulation:
Database: Ovid MEDLINE(R)
Search Strategy:

```
1  exp multiple sclerosis/dt, th, im (16589)
2  exp immunomodulation/ (229961)
3  exp immunosuppressive agents/ (235978)
4  exp immunologic techniques/ (1168203)
6  1 and 2 (1184)
7  1 and 3 (1831)
8  1 and 4 (2881)
9  6 or 7 or 8 (4864)
```

KQ1. MS/Drug Names:
Database: Ovid MEDLINE(R)
Search Strategy:

```
1  Multiple Sclerosis, Relapsing-Remitting/dt, im, th [Drug Therapy, Immunology, Therapy]
2  exp Interferon-beta/
3  interferon beta.mp.
4  glatiramer acetate.mp.
5  natalizumab.mp.
6  teriflunomide.mp.
7  3 or 4 or 5 or 6
```
8  1 and 7
9  limit 8 to (addresses or autobiography or bibliography or biography or classical article or
  comment or editorial or historical article or interactive tutorial or lectures or news or
  newspaper article or patient education handout)
10  8 not 9
11  limit 10 to (english language and humans and yr="1990 -Current")

For KQ2, we searched Medline via Ovid, Cochrane Libraries, PsychiInfo, and CINAHL,
modifying the Medline searches for the other databases.

KQ2. MS/Patient Preference:
Database: Ovid MEDLINE(R)
Search Strategy:

1  exp multiple sclerosis/dt, th, im
2  exp patient preference/
3  exp attitude to health/
4  exp physician-patient relations/
5  exp decision making/
6  exp choice behavior/
7  exp decision support techniques/
8  exp personal autonomy/
9  exp patient participation/
10 decision making.mp.
11 decision support.mp.
12 risk communication$.mp.
13 shared decision$.mp.
14 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
15 1 and 14
16 exp health knowledge, attitudes, practice/
17 exp *multiple sclerosis/
18 16 and 17
19 15 or 18
20 exp multiple sclerosis/px
21 14 and 20
22 19 or 21
Appendix B. Excluded Studies
(reason for exclusion appears at the end of each reference)


81. West TW, Cree BA. Natalizumab dosage suspension: are we helping or hurting? Annals of Neurology 2010 Sep; 68(3):395-9 (design).

Appendix C. Evidence Tables

Appendix Table C1. Evidence table for KQ1 (not natalizumab interruption or pregnancy) studies ................................................................. C-2
Appendix Table C2. Analytic set outcomes ......................................................................................................................... C-5
Appendix Table C3. Analytic set harms ......................................................................................................................... C-7
Appendix Table C4. Discontinuation details by study (if reported) from full reporting set ...................................................... C-10
Appendix Table C5. Studies of natalizumab discontinuation ................................................................................................. C-12
Appendix Table C6. Outcomes for studies of natalizumab discontinuation ............................................................................ C-14
Appendix Table C7. Studies of women with MS who experienced pregnancy and their fetuses ......................................................................................................................... C-16
Appendix Table C8. Fetal outcomes for DMT exposure and maternal outcomes for drug holiday ......................................................................................................................... C-18
References for Appendix C ............................................................................................................................................... C-23
### Appendix Table C1. Evidence table for KQ1 (not natalizumab interruption or pregnancy) studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Study Aim</th>
<th>Patient Population</th>
<th>Drugs</th>
<th>Study Design</th>
<th>Duration: Treatment, Followup</th>
<th>Comparison(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bencsik, 2006</td>
<td>Hungary</td>
<td>To evaluate long-term efficacy</td>
<td>36: 34 RRMS, 2 RPMS mean age 36.0</td>
<td>IFNβ-1b</td>
<td>Case series</td>
<td>Tx: 6y Fu: 6y</td>
<td>By duration of therapy</td>
</tr>
<tr>
<td>Bermel, 2013</td>
<td>United States</td>
<td>To evaluate long-term tolerability, efficacy</td>
<td>172 RRMS (122 living, 14 died, 24 unknown) 78% F mean age 35.8</td>
<td>IFNβ-1a</td>
<td>Open-label extension</td>
<td>Tx: med. 13.3y (mean 12.1, range 3–15) Fu: med. 16.3y</td>
<td>56 currently receiving IFNβ-1a im vs. 66 not; by EDSS quartile</td>
</tr>
<tr>
<td>Bergamaschi 2012</td>
<td>Italy</td>
<td>To investigate disability progression</td>
<td>1178 RRMS</td>
<td>Mixed IFNβ and glatiramer acetate</td>
<td>Tx: NR Fu: median 16.5y</td>
<td>Treated vs untreated by Bayesian risk quartile</td>
<td></td>
</tr>
<tr>
<td>Carmona, 2008</td>
<td>Spain</td>
<td>To investigate early prognostic markers, efficacy</td>
<td>115 RRMS 68% F</td>
<td>IFNβ-1b</td>
<td>Prospective cohort with historical control</td>
<td>Tx: NR Fu: mean 4.6y (range 0.5-7.1), med 5.6y</td>
<td>115 treated, 44 untreated historic cohort</td>
</tr>
<tr>
<td>Confavreux, 2012</td>
<td>North America and Europe</td>
<td>To evaluate tolerability, efficacy</td>
<td>147: 131 RRMS, 16 SPMS mean age 40</td>
<td>Teriflunomide</td>
<td>Open-label extension of RCT</td>
<td>Tx: mean 5.6y, med. 7.1 (SD 2.7, range 0.05-8.5) Fu: up to 8.5y</td>
<td>52 7mg 40 14mg 29 pl.+7mg 26 pl.+14mg</td>
</tr>
<tr>
<td>Cunningham, 2010</td>
<td>Sweden</td>
<td>To explore stopping, switching, continuing.</td>
<td>259 MS 71% F</td>
<td>IFNβ mixed</td>
<td>Cohort study</td>
<td>Tx: 3y+ continuers, 1.2y stoppers, 1.2y switchers Fu: NR</td>
<td>80 tx at least 3y, 38 stopped, 141 switched</td>
</tr>
<tr>
<td>Debouverie, 2007</td>
<td>France</td>
<td>To evaluate long-term safety, efficacy</td>
<td>205 RRMS 77% F mean age 38.5</td>
<td>GA</td>
<td>Case series</td>
<td>Tx: med. 5y Fu: 3.5-8y</td>
<td>By tx duration</td>
</tr>
<tr>
<td>Evans, 2012</td>
<td>Canada</td>
<td>To describe long-term DMT persistence</td>
<td>1896 MS 75% F mean age 42.5</td>
<td>Mixed</td>
<td>Cohort study</td>
<td>Tx: med. 2.9y initial DMT, med. 6.3y any first-line DMT Fu: up to 14y</td>
<td>By specific drug</td>
</tr>
<tr>
<td>Ford, 2010</td>
<td>United States</td>
<td>To evaluate long-term safety, efficacy</td>
<td>232 mITT, 100 ongoing cohort, mITT: 73% F mean age 35.5</td>
<td>GA</td>
<td>Cohort study</td>
<td>Tx: mITT mean 8.6y (SD 5.2), ongoing mean 13.6y (SD 1.3) Fu: up to 15y</td>
<td>232 mITT, 100 ongoing, 131 withdrawn</td>
</tr>
<tr>
<td>Goodin, 2012</td>
<td>11 North American centers</td>
<td>To investigate effect of early tx on survival</td>
<td>366 identified of 372 RRMS</td>
<td>IFNβ-1b</td>
<td>Long term followup</td>
<td>Tx: NR Fu: 21y</td>
<td>149 treated 123 placebo</td>
</tr>
<tr>
<td>Kappos, 2009</td>
<td>North America and Europe</td>
<td>To compare early vs. delayed tx</td>
<td>418 MS</td>
<td>IFNβ-1b</td>
<td>Open-label extension of RCT</td>
<td>Tx: med. 5y early tx, med. 3y delayed tx Fu: 5y</td>
<td>261 early tx, 157 delayed tx</td>
</tr>
<tr>
<td>Study Country</td>
<td>Study Aim</td>
<td>Patient Population</td>
<td>Drugs</td>
<td>Study Design</td>
<td>Duration: Treatment, Followup</td>
<td>Comparison(s)</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-------</td>
<td>----------------------------------------</td>
<td>---------------------------------------------------</td>
<td>------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Mesaros, Serbia</td>
<td>To assess frequency and reasons for stopping</td>
<td>290 RRMS 71% F mean age 38.0 EDSS at least 3.5</td>
<td>IFNβ-1a, IFNβ-1b Cohort study</td>
<td>Tx: IFNβ-1a mean 3.7y, IFNβ-1b mean 3.2y Fu: up to 6y, mean 3.5y</td>
<td>169 IFNβ-1a sc, 121 IFNβ-1b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Milanese, Italy</td>
<td>To compare long-term efficacy</td>
<td>294 RRMS</td>
<td>IFNβ-1a, IFNβ-1b, GA Case series 20/26 centers</td>
<td>Tx mean: IFNβ-1b 3.5y, IFNβ-1a 2.3-3.2y, GA 1.6y Fu: up to 5y</td>
<td>By specific drug</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller, United States</td>
<td>To evaluate long-term safety, efficacy</td>
<td>46 RRMS</td>
<td>GA</td>
<td>TX: med. 12y (range 1-22y) Fu: NR</td>
<td>18 continuing, 28 discontinued</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Rourke, Ireland</td>
<td>To investigate DMT discontinuation</td>
<td>394: 246 RRMS, 148 SPMS mean age 38</td>
<td>IFNβ mixed Cohort study</td>
<td>Tx: NR Fu: med. 4.1y (up to 8y)</td>
<td>246 RRMS, 148 SPMS; by specific drug, by reasons for discon.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Rourke, Ireland</td>
<td>To evaluate efficacy, prognostic markers</td>
<td>175 RRMS 72% F</td>
<td>IFNβ mixed Cohort study</td>
<td>Tx: &gt;2y Fu: mean 5.1y (range 2-10)</td>
<td>175 treated, 185 historical cohort (mean obs 3y)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patti, Italy</td>
<td>To evaluate, compare long term safety, efficacy</td>
<td>126 RRMS 59% F mean age 36.7</td>
<td>IFNβ-1a, IFNβ-1b Cohort study</td>
<td>Tx: 6y Fu: 6y</td>
<td>62 IFNβ1a im, 64 IFNβ1b sc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portaccio, Italy</td>
<td>To investigate DMT discontinuation</td>
<td>225 RRMS 70% F mean age 36.6</td>
<td>IFNβ mixed (2 of 4 IFN forms) Cohort study</td>
<td>Tx: mean 3.1y β-1a im, 3.3y β-1b Fu: mean 4.7y β-1a im, 6.0y β-1b</td>
<td>By specific drug, by reasons for discon.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rio, Spain</td>
<td>To assess safety, efficacy, investigate adherence to DMT</td>
<td>382 RRMS 632 MS: 134 SPMS</td>
<td>IFNβ-1a, IFNβ-1b Open-label nonrandomized post-marketing observational</td>
<td>Tx: NR Fu: mean 4y (range 2-8)</td>
<td>By reasons for discon. (drugs aggregated)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rio, Spain</td>
<td>To report post-marketing experience</td>
<td>146 SPMS 62% F mean age 45.1</td>
<td>IFNβ-1b Post-marketing study</td>
<td>Tx: NR Fu: med. 5y (range 1-9.6)</td>
<td>By clinical status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shirani, Canada</td>
<td>To investigate disability progression</td>
<td>2656 RRMS 76% F</td>
<td>IFNβ mixed Retrospective cohort with historical and contemporary control groups</td>
<td>Tx: med. 5.1y Fu: med 5.1y tx, 4.0y untreated, 10.8y historical</td>
<td>868 treated, 929 untreated, 959 historical</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C-3
<table>
<thead>
<tr>
<th>Study Country</th>
<th>Study Aim</th>
<th>Patient Population</th>
<th>Drugs</th>
<th>Study Design</th>
<th>Duration: Treatment, Followup</th>
<th>Comparison(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorensen, 2006&lt;sup&gt;61&lt;/sup&gt; Denmark</td>
<td>To provide data on DMT use in a population</td>
<td>2393 RRMS</td>
<td>IFNβ mixed &lt;br&gt; Cohort study</td>
<td>Tx mean: IFNβ-1b 3.8y, IFNβ-1a 2.7-3.7y, GA 1.8y &lt;br&gt; Fu: mean 3.7y (range 0-8.8)</td>
<td>By specific drug, by tx duration</td>
<td></td>
</tr>
<tr>
<td>Trojano, 2005&lt;sup&gt;62&lt;/sup&gt; Italy</td>
<td>Surveillance, efficacy</td>
<td>1163: 943 RRMS, 220 SPMS (not analyzed)</td>
<td>IFNβ mixed (n for drugs not same as analysis) &lt;br&gt; Post-marketing study</td>
<td>Tx: results for 4y &lt;br&gt; Fu: up to 6y</td>
<td>By specific drug &lt;br&gt; (visual images only)</td>
<td></td>
</tr>
<tr>
<td>Trojano, 2007&lt;sup&gt;63&lt;/sup&gt; Italy</td>
<td>To investigate disability progression</td>
<td>1504 RRMS &lt;br&gt; 69% F &lt;br&gt; mean age 33.6</td>
<td>IFNβ-1a, IFNβ-1b &lt;br&gt; Cohort study</td>
<td>Tx: NR &lt;br&gt; Fu: med. 5.7y (up to 7y)</td>
<td>1103 treated, 401 untreated</td>
<td></td>
</tr>
<tr>
<td>Trojano, 2009&lt;sup&gt;64&lt;/sup&gt; Italy</td>
<td>To evaluate the effectiveness of early treatment</td>
<td>2570 RRMS &lt;br&gt; 69% F &lt;br&gt; mean age 33.5</td>
<td>IFNβ mixed &lt;br&gt; Cohort study</td>
<td>Tx: NR &lt;br&gt; Fu: median 4.5y (up to 7y)</td>
<td>310 early tx, 2260 delayed tx</td>
<td></td>
</tr>
<tr>
<td>Tedeholm, 2013&lt;sup&gt;65&lt;/sup&gt; Sweden</td>
<td>To investigate disability progression</td>
<td>916 RRMS</td>
<td>Mixed IFNβ and glatiramer acetate</td>
<td>Tx: NR &lt;br&gt; Fu: 12y</td>
<td>186 untreated, 730 treated</td>
<td></td>
</tr>
<tr>
<td>Uitdehaag, 2011&lt;sup&gt;66,-68&lt;/sup&gt; 22 centers in Europe, Canada, Australia</td>
<td>To evaluate safety, efficacy</td>
<td>382 RRMS</td>
<td>IFNβ-1a &lt;br&gt; Retrospective long-term followup of open-label extension</td>
<td>Tx: mean 3y min quartile, 7.5y max quartile by exposure time &lt;br&gt; Fu: up to 8y</td>
<td>Quartiles by dose &amp; time &lt;br&gt; 96 min quartile, 95 max quartile</td>
<td></td>
</tr>
</tbody>
</table>

AEs=adverse events; ARR=Annualized relapse rate; discon=discontinuion; DMT=disease modifying treatment; EDSS=Expanded Disability Status Scale; F=female; Fu=followup; IFNβ=interferon beta; GA=glatiramer acetate; im=intramuscular; LTFU=long-term followup; med=median ; mITT=modified intention-to-treat; MS=multiple sclerosis; NR=not reported; RRMS=relapse remitting multiple sclerosis; SD=standard deviation; SPMS=secondary-progressive multiple sclerosis; Tx=treatment; y=year
### Appendix Table C2. Analytic set outcomes

<table>
<thead>
<tr>
<th>Study Country</th>
<th>Relapse: Mean ARR (SD)</th>
<th>Disability: Change in Mean EDSS (SD) from Baseline</th>
<th>Progression to SPMS:</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergamaschi, 2012* Italy</td>
<td></td>
<td></td>
<td>Bayesian modeling by quartile based on risk of progressing to SPMS. Highest quartile; treated 25.4%, untreated 64.4% (RR 0.23, 95% CI, 0.15 to 0.35) Lowest quartile (RR 0.27, 95% CI, 0.13 to 0.56)</td>
<td></td>
</tr>
<tr>
<td>Bencsik, 2006† Hungary</td>
<td>Decreased from 1.29 to 0.25 (p&lt;0.001) at 6y</td>
<td>Increased 0.5 (p=0.016) at 6y</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Carmona, 2008‡ Spain</td>
<td>Decreased from 3.2 at baseline to 1.7 at mean 4.6y, med 5.6y</td>
<td>NR</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Confavreux, 2012§ North America and Europe</td>
<td>Decreased from baseline in all four tx dose/crossover groups</td>
<td>Stable in all groups with no significant differences or changes</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Debouverie, 2007‖ France</td>
<td>Stable, 0.4–0.6 annually for 5y</td>
<td>Stable: 3.1 (1.7) at baseline to 3.3 (2.1) at med. 5y</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Ford, 2010¹⁰-¹⁴ United States</td>
<td>Decreased from 1.12 (0.82) at baseline to 0.25 (0.34) at mean 13.6y</td>
<td>Increased 0.6 (2.0) at mean 13.6y 57% stable or improved</td>
<td>mITT cohort 59 (25%) at mean 9.8y ongoing cohort 35 (35%) at 12.0y withdrawn cohort 24 (18.5%) at mean 6.4y</td>
<td>NR</td>
</tr>
<tr>
<td>Goodin, 2012¹⁵ Reder 2010¹５,¹６ 11 North American centers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kappos, 2009¹⁸ North America and Europe</td>
<td>No significant changes or differences</td>
<td>No significant changes or differences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller, 2008²¹ United States</td>
<td>Decreased from 2.9 (1.4) at baseline to 0.1 (0.2) at med 12y</td>
<td>Increased 0.9 (1.9) at med. 12y, 57% stable or improved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patti, 2006* Italy</td>
<td>NR</td>
<td>IFN-β1a: increased 1.03 (1.35) at 6y IFNβ-1b: increased 0.97 (1.47) at 6y (p=0.47)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C-5
<table>
<thead>
<tr>
<th>Study Country</th>
<th>Relapse: Mean ARR (SD)</th>
<th>Disability: Change in Mean EDSS (SD) from Baseline</th>
<th>Progression to SPMS:</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portaccio, 2008&lt;sup&gt;25&lt;/sup&gt; Italy</td>
<td>β-1a im: 1.6 (1.0) at baseline to 0.4 (0.8) at 3.1y β-1b: 1.1 (0.7) at baseline to 0.7 (1.2) at 3.3y</td>
<td>Reported no significant changes or differences</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Rio, 2007&lt;sup&gt;26&lt;/sup&gt; Spain</td>
<td>Decreased from 0.65 at baseline (2y prior: 1.3) 0.22 at 4y</td>
<td>NR</td>
<td>Included only people with SPMS</td>
<td>NR</td>
</tr>
<tr>
<td>Rio, 2005&lt;sup&gt;27,28&lt;/sup&gt; Spain</td>
<td>IFNβ-1b: 1.36 (0.7) at baseline to 0.38 (0.45) at 4y, IFNβ-1a im: 1.07 (0.5) to 0.33 (0.48), IFNβ-1a sc: 1.21 (0.5) to 0.41 (0.5)</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Shirani, 2012&lt;sup&gt;29&lt;/sup&gt; Canada</td>
<td>NR</td>
<td>NR</td>
<td>IFNβ not associated with reduced hazard of progression to EDSS 6 compared with contemporary or historical control, with multivariate analysis</td>
<td>NR</td>
</tr>
<tr>
<td>Tedeholm, 2013&lt;sup&gt;30&lt;/sup&gt; Sweden</td>
<td>NR</td>
<td>NR</td>
<td>No difference between treated and untreated.</td>
<td>NR</td>
</tr>
<tr>
<td>Uitdehaag, 2011&lt;sup&gt;31-35&lt;/sup&gt; PRISMS Study 22 centers in Europe, Canada, Australia</td>
<td>No statistical analysis. β-1a sc by time exposure: 0.76 (0.55) min time quartile, 0.51 (0.49) max time quartile At 7-8 y</td>
<td>No statistical analysis – lower percentage confirmed EDSS progression</td>
<td>No statistical analysis – lower percentage convert to SPMS</td>
<td>NR</td>
</tr>
</tbody>
</table>

ARR=Annualized relapse rate; EDSS=Expanded Disability Status Scale; Fu=followup; IFNβ=interferon beta; GA=glatiramer acetate; im=intramuscular; LTFU=long-term followup; med=median; mITT=modified intention-to-treat; NR=not reported; sc=subcutaneous; SD=standard deviation; Tx=treatment; y=year
## Appendix Table C3. Analytic set harms

<table>
<thead>
<tr>
<th>Author/Year</th>
<th>Harms/AEs Overall</th>
<th>Harms Information</th>
<th>Context/Comments</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confavreaux, 2012&lt;sup&gt;6&lt;/sup&gt; Teriflunomide N=147</td>
<td>Yes</td>
<td>AEs – full page table 2</td>
<td>TEAEs: treatment-emergent adverse effects (1&lt;sup&gt;st&lt;/sup&gt; dose to within 16 wks of last dose) Other AEs (spontaneously reported at visits)</td>
<td>Safety data (TEAEs) presented for (core + extension) and according to tx received during extension phase (placebo group was randomized to either dose for extension)</td>
</tr>
<tr>
<td>Debouverie, 2007&lt;sup&gt;st&lt;/sup&gt; GA: intolerant of INF N=205</td>
<td>Yes: AEs Table 6 &amp; text p 1271</td>
<td>AEs and discontinue for AE Table 6. AE and discontinue reasons not mutually exclusive</td>
<td>Excluded patients who stopped GA (40% of patients stopped GA before LT followup → excluded)</td>
<td>87.3% overall had AEs: (179/205) Local injection site rxns: 166 (81%) Systemic rxns: 101 (49.3%): immediate post-injection: flushing, chest tightness, shortness of breath, palpitations, anxiety. *Only 5% of patients overall discontinued because of these AEs</td>
</tr>
<tr>
<td>Ford, 2010&lt;sup&gt;10&lt;/sup&gt; LT GA Ford, 2006&lt;sup&gt;11&lt;/sup&gt; N=232</td>
<td>Yes but text only, pg 347, no N (%) given</td>
<td></td>
<td>1. Most commonly reported AEs: Accidental injury, muscle weakness, back pain, dizziness, depression, hypoesthesia, paresthesia, insomnia, URIs, UTIs, headaches, pain. 2. AEs thought to be related to GA: local injection-site rxns, IPIR (vasodilation, chest pain, palpitation, tachycardia, dyspnea. *No apparent time-dependent AEs emerged. No evidence of renal dysfunction, immunosuppression, malignancy, or development of other autoimmune disease was observed. One death (resp. failure during pneumonia)</td>
<td></td>
</tr>
<tr>
<td>Johnson, 2000&lt;sup&gt;12&lt;/sup&gt; (also Johnson 03,05)</td>
<td>Yes: 3 statements in text</td>
<td>AEs brief text p 260 Safety:</td>
<td>Can’t tell if % are for overall days of events or number of events</td>
<td>&quot;Most common reported AEs were injection site rxns. Group A (2.4%), Group B (0.9%). No reports of skin necrosis. No lab deviations&quot;</td>
</tr>
<tr>
<td>Gold, 2005&lt;sup&gt;36&lt;/sup&gt; INF-1a</td>
<td>Yes detailed: Table 2 and text p 651-654. Table 3 lab abnormal</td>
<td></td>
<td></td>
<td>Most common AEs in patients originally randomized to active tx: Injection-site inflammation (72%), headache (71%), flu-like symptoms (69%). All were generally mild and most common in first month of tx. *4yr AE profiles were comparable with those during the initial phase of study and for the most part, with each other. No association. between INF1a and depression/suicide. Lab: most common abnormal labs were asymptomatic lymphopenia and elevated serum liver transaminase levels (mild and resolved spontaneous) ** Serious AEs detailed in text p 654. Also deaths</td>
</tr>
<tr>
<td>Author/Year</td>
<td>Harms/AEs Overall</td>
<td>Harms Information</td>
<td>Context/Comments</td>
<td>Results</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>PRISMS, 2001&lt;sup&gt;16&lt;/sup&gt;</td>
<td>Yes p 1633 text; Table 4 p 1635.</td>
<td>Lists AEs by N - no %’s given - so need to find denominators in text or other tables</td>
<td>AEs similar to those observed in PRISMS-2; most mild Table 4: AEs by: (yrs1-4) and (yrs 3-4); Table listed by cases and without percentages. Most common: injection site, flu-like symptoms. Also lists lab abnormalities text p1633 &amp; Table 4</td>
<td></td>
</tr>
<tr>
<td>(same sample as Gold, Utidehaag</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goodin, 2012&lt;sup&gt;6&lt;/sup&gt; Reder, 2010&lt;sup&gt;16&lt;/sup&gt;</td>
<td>Yes: AEs Table 2 and mortality</td>
<td>“The great number of tx sequences renders individual analyses and commentary difficult, but some observed AEs are likely derived from tx with other agents and not from INFB” p1878. Detailed death info by dose, NAb status p1881</td>
<td>AEs given for patients continuously using INFB-1b in the 2 yrs prior to LTFU (n=69) Of n=69 patients at LTF visit: Table 2 Injection site rxn 81%; Depression 42%; Flu like symptoms: 32%; Headache 28%; Malaise 23%; Fever 22%; Myalgia 22%; Liver transaminase increase 10%.</td>
<td></td>
</tr>
<tr>
<td>INFb-1b N=328</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kappos, 2009&lt;sup&gt;18&lt;/sup&gt;</td>
<td>Yes –text p 991: web appendix</td>
<td>Comparison is early vs. delayed tx groups INFB-1b.</td>
<td>web appendix for table.</td>
<td>% of patients who had ‘at least one serious AE: Early tx: 61 (21%); Delayed tx: 42 (24%). 40 (23%) of 173 delayed who had post base blood taken had at least 1 + for NAb. More on titers p 992</td>
</tr>
<tr>
<td>N=418</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller, 2008&lt;sup&gt;21&lt;/sup&gt;</td>
<td>Yes text only p 497</td>
<td>No. discontinuing due to serious adverse events.</td>
<td>“Most common AEs in ≥50% of patients:” Injection site rxns (soreness, redness, swelling, itching). 6 who took GA for up to 22 yrs reported lipoatrophy. None reported skin necrosis</td>
<td></td>
</tr>
<tr>
<td>22 yrs N=46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patti, 2006&lt;sup&gt;24&lt;/sup&gt;</td>
<td>Text only p 244. Also NAbs</td>
<td>6 yr AE: text summary only no N or % Number with NAbs at yr 6: text p 244</td>
<td>(Also text AEs in discontinues: authors mix AE and AE/discontinue info in both subsections)</td>
<td>“Most frequent SEs were flu like symptoms, fever, headache, injection-site rxn, fatigue, myalgia, increased spasticity and depression. Headache was significantly more common in Group A (Avonex) and injection site rxn in Group B Betaferon. NSD between groups in other SEs. NAbs yr 6: text p 244</td>
</tr>
<tr>
<td>INFb1a vs INFb1b, 6 yrs of therapy N=126</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portaccio, 2008&lt;sup&gt;25&lt;/sup&gt;</td>
<td>Yes – minimal - text p 132-133 and Table 2</td>
<td>FU duration mean 4.2 yrs. No. discontinuing due to serious adverse events</td>
<td>Patients discontinuing due to SE (14.7%) did so significantly earlier than patients who suspended IFNB due to perceived lack of efficacy (28.9%). Med survival times 1.0y vs 2.3y (log rank test p=0.009). Suspension due to SE associated with a higher relapse rate in yr prior to IFNB. Confirmed in patients with followup of at least 4 yr.</td>
<td></td>
</tr>
<tr>
<td>IFN mixed N=225</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rio, 2005&lt;sup&gt;27,28&lt;/sup&gt;</td>
<td>Yes – little</td>
<td>Summary comments in text p 798 Safety. Minimal details</td>
<td>“No unexpected AEs. All treatments were well tolerated and AEs in accordance with those previously associated with IFNB (data not shown) Most AEs were more frequent in first months of treatment and reduced after first 6 mo. Frequency of AEs was similar in all groups except for one AE: injection site reactions were more common in SQ INFB. On the other hand, 3 patients developed allergic reaction to Rebif with urticarial and angioedema. In all 3 INFB was interrupted and symptoms resolved with specific treatment”</td>
<td></td>
</tr>
<tr>
<td>Author/Year</td>
<td>Harms/AEs Overall</td>
<td>Harms Information</td>
<td>Context/Comments</td>
<td>Results</td>
</tr>
<tr>
<td>------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>---------</td>
</tr>
</tbody>
</table>
| Rio, 2007**  
N=146 SPMS 1y,89  
at least 3y | Yes  
minimal -  
not useable;  
See text | Text comments – no denominators p 852 | Unclear what AE text is for stoppers vs. nonstoppers, nor what the denominator is | “…AEs were in accordance with those previously associated with INFβ (data not shown). Most AEs were more frequent during the first months of treatment then reducing in frequency after the first 6 months. Only in 3 patients did AEs lead to therapy discontinuation. Nevertheless, some unexpected AEs were observed as follows: 4 patient died (3 sepsis, 1 pulmonary hemorrhage), 1 patient had herpes zoster, 1 intracerebral hemorrhage and one had GI bleeding” p 852 |

ARR=Annualized relapse rate; EDSS-Expanded Disability Status Scale; Fu=followup; IFNβ=interferon beta; GA=glatiramer acetate; im=intramuscular; LTFU=long-term followup; med=median; mITT=modified intention-to-treat; NR=not reported; sc=subcutaneous; SD=standard deviation; Tx=treatment; y=year
<table>
<thead>
<tr>
<th>Author, Year Drugs</th>
<th>Adverse Event</th>
<th>Lack of Efficacy</th>
<th>Progression of Disability</th>
<th>Intended Pregnancy</th>
<th>Long-term Stable MS</th>
<th>Death</th>
<th>Protocol Violation</th>
<th>Other Treatment</th>
<th>Patient Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Glatiramer acetate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debouverie, 2007</td>
<td>Y</td>
<td>Y “aggravation of MS”</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y “other”*</td>
</tr>
<tr>
<td>Ford, 2010</td>
<td>Y</td>
<td>Y “perception of disease worsening”</td>
<td>Y</td>
<td>Y</td>
<td>Y “benign”</td>
<td>Y</td>
<td>Y “difficulty, inability, or unwillingness to adhere to protocol”</td>
<td>Y</td>
<td>Y 95/132 “patient decision” with subcategories</td>
</tr>
<tr>
<td>Miller, 2008</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y “did not complete or return forms”</td>
<td>Y</td>
<td>Y 15/28 “patient withdrew consent” 8/15 progressed at least 1.0 EDSS</td>
</tr>
<tr>
<td><strong>Teriflunomide</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confavreux, 2012</td>
<td>Y</td>
<td>Y “perception of lack of efficacy”</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>“subject did not wish to continue”, “other”</td>
</tr>
<tr>
<td><strong>Teriflunomide</strong></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>INFβ-1a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uitdehaag, 2011** Gold 2005**</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>“Y “patient decision”, “other”</td>
</tr>
<tr>
<td><strong>Interferon beta-1a</strong></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>INFβ-1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benckisik, 2006</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Carmona, 2008</td>
<td>Y</td>
<td>Y “insufficient efficacy”</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y “voluntary withdrawal”**</td>
</tr>
<tr>
<td>Kappos, 2009*</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y “other”**, “withdrew consent”*</td>
<td></td>
</tr>
<tr>
<td>Rio, 2007**</td>
<td>Y</td>
<td>Y “inefficacy (increase of disability or rr)”</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y “by own will”**</td>
</tr>
<tr>
<td><strong>Interferon beta-1b</strong></td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>INFβ mixed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cunningham, 2010</td>
<td>Y</td>
<td>Y “perceived lack of”</td>
<td>N</td>
<td>N (excluded)</td>
<td>Y “long-term”</td>
<td>NA</td>
<td>NA</td>
<td>Y</td>
<td>Y “miscellaneous”***</td>
</tr>
<tr>
<td>Author, Year</td>
<td>Drugs</td>
<td>Adverse Event</td>
<td>Lack of Efficacy</td>
<td>Progression of Disability</td>
<td>Intended Pregnancy</td>
<td>Long-term Stable MS</td>
<td>Death</td>
<td>Protocol Violation</td>
<td>Other Treatment</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>--------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>-------</td>
<td>-------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Mesaros, 2012</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>O'Rourke, 2007</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>O'Rourke, 2005</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Patti, 2005</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Portaccio, 2008</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Troiano, 2005</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Interferon beta mixed</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Milanese, 2005</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Rio, 2005</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Sorenson, 2006</td>
<td>Interferon beta mixed</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>DMT mixed</td>
<td>DMT mixed</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

DMT=disease-modifying therapy; *not explained; **explained further
### Appendix Table C5. Studies of natalizumab discontinuation

<table>
<thead>
<tr>
<th>Study, Location</th>
<th>Population, Mean EDSS</th>
<th>Design</th>
<th>Sample Size, Reason(s) for Interruption</th>
<th>DMT Prior to NTZ</th>
<th>Duration of NTZ</th>
<th>Duration of Interruption</th>
<th>DMT During Interruption</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ARR Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Havla, 2013&lt;sup&gt;39&lt;/sup&gt; Germany</td>
<td>MS patients</td>
<td>Retrospective cohort</td>
<td>n=36 26/36 fingolimod: 25 fear of PML, 2 treatment failure, 4 desire to switch; 10/36 no DMT: 5 fear of PML, 3 family planning, 1 treatment failure, 1 side effects</td>
<td>All, NR details</td>
<td>median 27.6 mo (median 27 doses, range 6-57)</td>
<td>median 13 mo observation; median 13.7 wks (fingolimod) or no DMT</td>
<td>26/36 FD, 10/36 none</td>
</tr>
<tr>
<td>Havla, 2011&lt;sup&gt;40&lt;/sup&gt; Germany</td>
<td>MS patients, EDSS 3.7</td>
<td>Prospective cohort</td>
<td>n=13 Mostly drug holiday (&quot;mainly because of fear of PML&quot;)</td>
<td>All, NR details</td>
<td>mean 26.3 mo (range 3–45)</td>
<td>mean 8.7 mo (range 3-16)</td>
<td>7/13 GA, 6/13 none</td>
</tr>
<tr>
<td>Jokubaitis, 2014&lt;sup&gt;41&lt;/sup&gt; Australia and International</td>
<td>MS patients</td>
<td>Retrospective cohort</td>
<td>N=89/536 NR</td>
<td>89 NTZ to FD 350 IFNβ or GA to FD 97 none to FD</td>
<td>31.8 mo</td>
<td>median 2.6 mo for NTZ to FD</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Kaufman, 2011&lt;sup&gt;42&lt;/sup&gt; United States</td>
<td>RRMS patients, EDSS 2.4</td>
<td>Retrospective chart review</td>
<td>n=48 Drug holiday or NR (advised drug holiday after 12 mo on NTZ)</td>
<td>39 failing DMT, 2 none, 7 NR</td>
<td>mean 12.3 mo (range 4–24)</td>
<td>mean 11.3 mo (range 7–24)</td>
<td>23 GA, 13 IFNβ, 7 other, 5 none; DMT initiated 3 mo after NTZ or without pause</td>
</tr>
<tr>
<td>Kerbrat, 2011&lt;sup&gt;43&lt;/sup&gt; France</td>
<td>RRMS patients with very active disease, EDSS 2.7</td>
<td>Retrospective cohort</td>
<td>n=27 17/27 drug holiday, 4/27 side effects (frequent infections: 2, rash: 1, fatigue: 1), 3/27 pregnancy planning, 3/27 patient stopped</td>
<td>All, NR details</td>
<td>mean 12 mo (range 6-23)</td>
<td>mean 6 mo</td>
<td>None</td>
</tr>
<tr>
<td>O’Connor, 2011&lt;sup&gt;44&lt;/sup&gt; Canada</td>
<td>MS patients, EDSS 2.4</td>
<td>Prospective, post hoc analysis of trial data</td>
<td>n=949 Drug holiday</td>
<td>NR</td>
<td>28.4 mo, median 34 doses (range 1-41)</td>
<td>8 mo observation</td>
<td>9.9% IFNβ, 2.4% GA,</td>
</tr>
<tr>
<td>Rossi, 2013&lt;sup&gt;45&lt;/sup&gt; Italy</td>
<td>RRMS patients, EDSS 2.8</td>
<td>Prospective cohort</td>
<td>n=40 Drug holiday</td>
<td>35 IFNβ, 5 none</td>
<td>mean 15.2 mo (range 12-18)</td>
<td>up to 12 mo</td>
<td>GA 4 wks after NTZ</td>
</tr>
<tr>
<td>Study, Location</td>
<td>Population, Mean EDSS</td>
<td>Design</td>
<td>Sample Size, Reason(s) for Interruption</td>
<td>DMT Prior to NTZ</td>
<td>Duration of NTZ</td>
<td>Duration of Interruption</td>
<td>DMT During Interruption</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------</td>
<td>--------</td>
<td>----------------------------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Sorensen, 2014, Denmark</td>
<td>MS patients with very active disease, EDSS 4.1 (range 0-8)</td>
<td>Retrospective; national registry</td>
<td>n=375 186 preferred to switch (mostly due to fear of PML), 44 treatment failure, 43 neutralizing antibodies, 26 pregnancy related, 20 side effects, 48 other, 8 unknown</td>
<td>320 first-line DMT, 44 mitoxantrone, 9 none, 2 IVIG</td>
<td>mean 32.4 mo (range 5.8–75.6 mo)</td>
<td>mean 3.8 mo (range 0–53.3 mo)</td>
<td>244 FD, 36 mitoxantrone, 30 resumed NTZ, 15 GA, 14 IFNβ s.c., 9 other IFNβ, 17 other, 10 none</td>
</tr>
<tr>
<td>MRI Outcomes only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boriello, 2012 Spain</td>
<td>RRMS patients, median EDSS 5.0</td>
<td>Prospective cohort post-marketing</td>
<td>n=23 Drug holiday</td>
<td>NR</td>
<td>mean 19 doses (range 12–28)</td>
<td>mean 117 days (90-150)</td>
<td>Pulsed steroids</td>
</tr>
<tr>
<td>Boriello, 2011 Spain</td>
<td>MS patients</td>
<td>Prospective cohort post-marketing</td>
<td>n=21 Drug holiday</td>
<td>1 none, 2 NR, 18 DMT NR details</td>
<td>24 consecutive mo</td>
<td>mean 111.5 days (range 90-174)</td>
<td>None</td>
</tr>
<tr>
<td>Miravalle, 2011 United States</td>
<td>MS patients with active disease, 75% RRMS, 25% SPMS, EDSS 3.3</td>
<td>Prospective cohort</td>
<td>n=32 Drug holiday</td>
<td>All, NR details</td>
<td>mean 17.3 mo (range 12+)</td>
<td>mean 4 mo</td>
<td>None</td>
</tr>
</tbody>
</table>

**DMT**=disease-modifying treatment; **EDSS**-Expanded Disability Status Scale; **Fu**=followup; **GA**=glatiramer acetate; **IFNβ**=interferon beta; **IVIG**=intravenous immunoglobulin; **mo**=month; **MS**=multiple sclerosis; **NR**=not reported; **NTZ**=natalizumab; **PML**=progressive multifocal leukoencephalopathy; **RRMS**=relapse remitting multiple sclerosis; **SPMS**=secondary progressive multiple sclerosis; **wks**=weeks
### Appendix Table C6. Outcomes for studies of natalizumab discontinuation

<table>
<thead>
<tr>
<th>Study, Location</th>
<th>ARR Before NTZ Mean (SD)</th>
<th>ARR After NTZ Mean (SD)</th>
<th>MRI Pre-NTZ Mean (SD) or Med (Range)</th>
<th>MRI After NTZ Mean (SD) or Med (Range)</th>
<th>Reported Rebound</th>
<th>Definition of Rebound/Severe Flare</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ARR and MRI outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Havla, 2013** Germany</td>
<td>fingolimod: 2.2 no DMT: 1.5</td>
<td>fingolimod: 0.0 no DMT: 1.5</td>
<td>(visual image only)</td>
<td>no significant change from baseline to post (visual image only)</td>
<td>3/36</td>
<td>Sustained EDSS worsening (&gt;1 EDSS steps) and widespread disease activity (&gt;1 GELs) on MRI</td>
</tr>
<tr>
<td>Havla, 2011** Germany</td>
<td>2.2 (1.5)</td>
<td>1.7 (1.4)</td>
<td>9/13 had GELs</td>
<td>9/12 had GELs</td>
<td>4/13</td>
<td>Subjective: severe relapse with sustained EDSS worsening and widespread disease activity on MRI</td>
</tr>
<tr>
<td>Jokubaitis, 2014** Australia and International</td>
<td>1.54</td>
<td>0.38</td>
<td>NR</td>
<td>NR</td>
<td>No</td>
<td>Compared ARR pre- and post-NTZ; also compared ARR and distribution of severe relapse among those who switched to FD from NTZ, first-line or no DMT</td>
</tr>
<tr>
<td>O’Connor, 2011** Canada</td>
<td>On-study placebo ARR: 0.73 at 7 mo (peak): 0.71</td>
<td>n=339: 1.6 (0.2) GELs</td>
<td>n=60 at 6 mo (peak observed): 1.2 (0.4) GELs</td>
<td>None</td>
<td>Worsening of disease activity beyond pretreatment levels</td>
<td></td>
</tr>
<tr>
<td>Kaufman, 2011** United States</td>
<td>0.52 (range 0–3)*</td>
<td>0.35</td>
<td>NR</td>
<td>NR</td>
<td>None</td>
<td>Subjective: severe relapse as has been reported by others (West and Cree 2010, subjective: clinically severe flare and mean 16 GELs (range 6-40))</td>
</tr>
<tr>
<td>Kerbrat, 2011** France</td>
<td>2.3 (1.1) at 6 mo: 1.8</td>
<td>21/27 had GELs, mean 2.8 (range 0-13)</td>
<td>13/19 had GELs, mean 9.1 (range 0-50)</td>
<td>4/27</td>
<td>Subjective: severe relapse and 20+ GELs (range 23-50)</td>
<td></td>
</tr>
<tr>
<td>Sorensen, 2014,** Denmark</td>
<td>0.94 1-6 mo: 0.63 7-12 mo: 0.55</td>
<td>NR</td>
<td>NR</td>
<td>83/375 or 42/375</td>
<td>Higher individual relapse rate after cessation of NTZ than before NTZ</td>
<td></td>
</tr>
<tr>
<td><strong>MRI outcomes only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rossi, 2013** Italy</td>
<td>2.3 (0.9)</td>
<td>0.6 (0.8)</td>
<td>Significantly higher (visual only)</td>
<td>15/29 had GELs, of the 15: mean 1.0 (1.1)</td>
<td>None</td>
<td>Subjective: MRI parameters were not consistent with rebound of disease activity</td>
</tr>
<tr>
<td>Boriello, 2012** Spain</td>
<td>2.1</td>
<td>NR</td>
<td>16/23 had CELs</td>
<td>7/23 had CELs of the 7, no med. change from baseline (2, 0-15) to post (3, 1-11)</td>
<td>2</td>
<td>Subjective: severe disabling relapse and a large number of CELs</td>
</tr>
<tr>
<td>Study, Location</td>
<td>ARR Before NTZ Mean (SD)</td>
<td>ARR After NTZ Mean (SD)</td>
<td>MRI Pre-NTZ Mean (SD) or Med (Range)</td>
<td>MRI After NTZ Mean (SD) or Med (Range)</td>
<td>Reported Rebound</td>
<td>Definition of Rebound/Severe Flare</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------</td>
<td>-------------------------</td>
<td>--------------------------------------</td>
<td>----------------------------------------</td>
<td>-----------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Boriello, 2011&lt;sup&gt;8&lt;/sup&gt; Spain</td>
<td>2.3</td>
<td>NR</td>
<td>16/19 had CELs</td>
<td>8/19 had CELs of the 9, no med. change from baseline (1, 0-12) to post (3, 0-20)</td>
<td>None</td>
<td>Subjective: disease activity worsens beyond pre-treatment severity</td>
</tr>
<tr>
<td>Miravalle, 2011&lt;sup&gt;3&lt;/sup&gt; United States</td>
<td>1y prior: 1.3 (1.1), 2y prior: 2.3 (1.5)</td>
<td>NR</td>
<td>n=31: 17/31 had GELs, mean 1.9 (2.9)</td>
<td>n=29: 12/29 had GELs, of the 12: mean 9.5 (12.4), (baseline mean 2.0 (2.8*), p .001</td>
<td>NR or 1?</td>
<td>Subjective: NR for observational study design</td>
</tr>
</tbody>
</table>

ARR=Annualized relapse rate; CELs=contrast-enhancing lesions; GELs=gadolinium enhancing lesions; med.=median; MRI=magnetic resonance imaging; mo=month; NR=not reported; NTZ=natalizumab. *This study confirmed ARR through clinical visits
<table>
<thead>
<tr>
<th>Study, Location</th>
<th>Population</th>
<th>Design</th>
<th>Number of Exposed Pregnancies, Intervention</th>
<th>Definition of Exposure, Mean Exposure Duration</th>
<th>Number of Unexposed Pregnancies, Comparison Group(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interferon beta 1a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandberg-Wollheim, 2011&lt;sup&gt;50&lt;/sup&gt; Sweden</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Prospective, global drug safety database</td>
<td>425 IFNβ-1a s.c.</td>
<td>During pregnancy 4 wks</td>
<td>None; general population data used for comparison</td>
</tr>
<tr>
<td>Sandberg-Wollheim, 2005&lt;sup&gt;51&lt;/sup&gt; Sweden</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Prospective, individual patient data from 8 trials</td>
<td>41 IFNβ-1a</td>
<td>Within 2 wks of conception or during pregnancy NR</td>
<td>6 Placebo 22 No DMT at least 2 wks prior to pregnancy</td>
</tr>
<tr>
<td><strong>Interferon beta mixed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amato, 2010&lt;sup&gt;52&lt;/sup&gt; Italy</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Prospective cohort, 21 clinics</td>
<td>88 IFNβ</td>
<td>Within 4 wks from conception or during pregnancy 4.6 wks</td>
<td>318 no DMT at least 4 wks prior to pregnancy</td>
</tr>
<tr>
<td>Boskovic, 2005&lt;sup&gt;53&lt;/sup&gt; Canada</td>
<td>Women with mostly* MS who experienced pregnancy and their fetuses</td>
<td>Prospective cohort, women who contacted risk counseling program</td>
<td>23 IFNβ</td>
<td>During pregnancy 9 wks (range 2-38) 21 MS: &lt;1&lt;sup&gt;st&lt;/sup&gt; tri 21 non-MS: 21, 38 wks</td>
<td>21 No DMT at least 4 wks prior to pregnancy 20 Healthy women who had been counseled for nausea or vomiting during pregnancy</td>
</tr>
<tr>
<td>Patti, 2008&lt;sup&gt;54&lt;/sup&gt; Italy</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Retrospective cohort, one clinic</td>
<td>14 IFNβ</td>
<td>During pregnancy 9 wks</td>
<td>17 DMT naive 7 No DMT at least 1 mo prior to pregnancy</td>
</tr>
<tr>
<td><strong>Interferon beta mixed, glatiramer acetate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragoso, 2013&lt;sup&gt;55&lt;/sup&gt; Brazil, UK, Mexico, Argentina</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Retrospective, voluntary database</td>
<td>41 GA 17 IFNβ 3 other DMT</td>
<td>Continuous DMT at least 8 wks during pregnancy 18.4 wks (range 8-40)</td>
<td>89 No DMT at least 3 mo prior to pregnancy</td>
</tr>
<tr>
<td>Hellwig, 2012&lt;sup&gt;56&lt;/sup&gt; Germany</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Retrospective, risk counseling database</td>
<td>78 IFNβ 15 IFNβ-1b 63 IFNβ-1a 41 GA</td>
<td>Last injection after last menstrual period 8.8 wks (5 DMT full term)</td>
<td>216 Last injection before last menstrual period</td>
</tr>
<tr>
<td>Finkelsztejn, 2011&lt;sup&gt;57&lt;/sup&gt; Brazil</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Retrospective cohort, provincial database</td>
<td>69 IFNβ 20 GA 10 other DMT</td>
<td>During pregnancy 8 wks “around 2 mo”</td>
<td>43 During pregnancy</td>
</tr>
<tr>
<td>Study, Location</td>
<td>Population</td>
<td>Design</td>
<td>Number of Exposed Pregnancies, Intervention</td>
<td>Definition of Exposure, Mean Exposure Duration</td>
<td>Number of Unexposed Pregnancies, Comparison Group(s)</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Lu, 2012&lt;sup&gt;34&lt;/sup&gt;</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Retrospective, voluntary database, selectively assembled by clinicians</td>
<td>15 IFNβ 6 GA</td>
<td>Within 1 mo of conception or during pregnancy &lt;2 mo</td>
<td>317 DMT naïve 80 No DMT at least 1 mo prior to pregnancy</td>
</tr>
<tr>
<td>Weber-Schoendorfer, 2009&lt;sup&gt;39&lt;/sup&gt;</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Prospective cohort, risk assessment program</td>
<td>69 IFNβ 31 GA</td>
<td>During pregnancy 6.9 wks GA 8.8 wks IFNβ</td>
<td>64 DMT naïve 1557 healthy women who had been counseled during pregnancy after nonteratogenic exposures</td>
</tr>
<tr>
<td>Giannini, 2012&lt;sup&gt;20&lt;/sup&gt;</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Prospective cohort, 21 clinics</td>
<td>17 GA</td>
<td>Within 4 wks from conception or during pregnancy 4.9 wks</td>
<td>318 No DMT at least 4 wks prior to pregnancy</td>
</tr>
<tr>
<td>Hellwig, 2011&lt;sup&gt;51&lt;/sup&gt;</td>
<td>Women with MS who experienced pregnancy and their fetuses</td>
<td>Prospective, risk counseling database</td>
<td>35 natalizumab</td>
<td>Within 8 wks of last menses or during pregnancy</td>
<td>23 DMT naïve</td>
</tr>
</tbody>
</table>

DMT=disease modifying treatment; IFNβ=interferon; GA=glatiramer acetate; MS=multiple sclerosis; wks=weeks

*14/16 women received IFN for MS, 1/16 for thrombocytosis, 1/16 for essential thrombocytthemia; in disease control group 10/12 women received IFN for MS, 1/12 for hepatitis C, 1/12 for wart on foot
<table>
<thead>
<tr>
<th>Author, Year, Location, Risk of Bias</th>
<th>Spontaneous Abortion (SA)</th>
<th>Fetal Death (FD)</th>
<th>Preterm Delivery (PD)</th>
<th>Fetal Outcomes: Reported Findings</th>
<th>Maternal Outcome: Postpartum Relapse Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interferon</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandberg-Wollheim, 2011(^{50})</td>
<td>e: 49/425</td>
<td>e: 6/425</td>
<td>NR</td>
<td>No significant differences in rates of SA between exposed and general population</td>
<td>NR</td>
</tr>
<tr>
<td>Amato, 2010(^{26}) Italy High</td>
<td>e: 7/88</td>
<td>c: 20/318</td>
<td>e: 1/88</td>
<td>Exposure associated with preterm delivery (propensity score-adjusted OR), not significantly associated with SA or FD.</td>
<td>NR</td>
</tr>
<tr>
<td>Patti, 2008(^{24}) Italy High</td>
<td>e: 0/14</td>
<td></td>
<td>e: 1/14</td>
<td>No significant differences in SA rates between exposed and unexposed groups.</td>
<td>NR</td>
</tr>
<tr>
<td>Boskovic, 2005(^{22}) Canada High</td>
<td>e: 9/23 c: 4/21</td>
<td>e: 2/23</td>
<td>c: 3/21</td>
<td>Exposed had higher rates of SA; no significant differences in rates of FD, PF.</td>
<td>NR</td>
</tr>
<tr>
<td>Sandberg-Wollheim, 2005(^{51})</td>
<td>e: 8/41 c: 0/28</td>
<td>e: 1/41</td>
<td>c: 0/28</td>
<td>No significant differences in rates of SA, FD, or PF between exposed and general population</td>
<td>NR</td>
</tr>
<tr>
<td>Fragoso, 2013(^{30}) Brazil, UK, Mexico, Argentina High</td>
<td>e: 0/17 c: 2/89</td>
<td>e: 0/17</td>
<td>c: 0/89</td>
<td>No significant differences in rates of SA, FD, or PF between exposed and unexposed</td>
<td>Women receiving any DMT at least 8 wks during pregnancy had significantly lower RR (p=0.001) following delivery compared to unexposed; no significant differences in RR 1y prior to or during pregnancy*</td>
</tr>
<tr>
<td>Finkelsztejn, 2011(^{27}) Brazil High</td>
<td>e: 0/69 c: 0/43</td>
<td>e: 0/69</td>
<td>c: 0/43</td>
<td>No significant differences in rates of SA or FD between exposed and unexposed</td>
<td>NR (descriptive only)</td>
</tr>
<tr>
<td>Lu, 2012(^{28}) Canada High</td>
<td>NR</td>
<td>e: 0/15</td>
<td>c: 1/397</td>
<td>No significant difference in rates of FD between exposed and unexposed</td>
<td>NR</td>
</tr>
<tr>
<td>Hellwig, 2012(^{30}) Germany High</td>
<td>NR</td>
<td>NR</td>
<td>e: 38.9 ± 2.4wks c: 39.1 ± 2.3wks</td>
<td>No significant differences in wks of gestation between exposed and unexposed</td>
<td>Exposed group had lower postpartum RR (p&lt;0.05)</td>
</tr>
<tr>
<td>Weber-Schoendorfer, 2009(^{39}) Germany High</td>
<td>e: 7/60 c: 6/61</td>
<td>e: 0/53</td>
<td>c: 0/57</td>
<td>No significant differences in rates of SA, FD, or PD between IFN-exposed and unexposed.</td>
<td>NR</td>
</tr>
<tr>
<td>Author, Year, Location, Risk of Bias</td>
<td>Spontaneous Abortion (SA)</td>
<td>Fetal Death (FD)</td>
<td>Preterm Delivery (PD)</td>
<td>Fetal Outcomes: Reported Findings</td>
<td>Maternal Outcome: Postpartum Relapse Rate</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td><strong>Glatiramer Acetate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragoso, 2013&lt;sup&gt;26&lt;/sup&gt; Brazil, UK, Mexico, Argentina High</td>
<td>e: 2/41 c: 2/89</td>
<td>e: 1/41 c: 0/89</td>
<td>e: 1/41 c: 1/89</td>
<td>No significant differences in rates of SA, FD, or PD between exposed and unexposed</td>
<td>Women receiving any DMT at least 8 wks during pregnancy had significantly lower RR (p=0.001) following delivery compared to unexposed; no significant differences in RR 1y prior to or during pregnancy *</td>
</tr>
<tr>
<td>Lu, 2012&lt;sup&gt;26&lt;/sup&gt; Canada High</td>
<td>NR</td>
<td>e: 0/6 c: 1/397</td>
<td>NR</td>
<td>No significant difference in rates of FD between exposed and unexposed</td>
<td>NR</td>
</tr>
<tr>
<td>Hellwig, 2012&lt;sup&gt;26&lt;/sup&gt; Germany High</td>
<td>NR</td>
<td>e: 39.2 ± 1.7wks c: 39.1 ± 2.3wks</td>
<td>No significant differences in wks of gestation between exposed and unexposed</td>
<td>Exposed group had lower postpartum RR (p&lt;0.05)</td>
<td></td>
</tr>
<tr>
<td>Giannini, 2012&lt;sup&gt;26&lt;/sup&gt; Italy</td>
<td>e: 1/17 c: 20/318</td>
<td>e: 0/17 c: 3/318</td>
<td>e: 4/17 c: 58/295</td>
<td>No significant differences in rates of SA, FD, or PD between exposed and unexposed</td>
<td>NR</td>
</tr>
<tr>
<td>Finkelsztejn, 2011&lt;sup&gt;27&lt;/sup&gt; Brazil High</td>
<td>e:0/20 c: 0/43</td>
<td>e: 0/20 c: 0/43</td>
<td>NR</td>
<td>No significant differences in rates of SA or FD between exposed and unexposed</td>
<td>NR (descriptive only)</td>
</tr>
<tr>
<td>Weber-Schoendorfer, 2009&lt;sup&gt;28&lt;/sup&gt; Germany High</td>
<td>e: 1/26 c: 6/61</td>
<td>e: 0/26 c: 0/57</td>
<td>e: 1/25 c: 8/55</td>
<td>No significant differences in rates of SA, FD, or PD between exposed and unexposed</td>
<td>NR</td>
</tr>
<tr>
<td><strong>Natalizumab</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hellwig 2011&lt;sup&gt;26&lt;/sup&gt; Germany High</td>
<td>e: 5/35 c: 1/23</td>
<td>e: 0/35 c: 1/23</td>
<td>e: 2/35 c: 1/23</td>
<td>No significant differences in rates of SA, FD, or PD between exposed and unexposed</td>
<td>No significant differences in RR between groups during pregnancy or post-partum; trend toward fewer post-partum relapses among exposed (e: 6/35, c: 9/23).</td>
</tr>
</tbody>
</table>

* timing for calculation of relapse rates not clearly reported  
DMT = disease modifying treatment; IFNβ = interferon beta; FD = fetal death; OR = odds ratio; PD = preterm delivery; SA = spontaneous abortion, e = exposed group, c = comparison group
References for Appendix C


Appendix D. Risk of Bias for Included KQ1 Studies

Appendix Table D1. KQ1 risk of bias for articles in analytic study set ........................................ D-2
Appendix Table D2. KQ1 risk of bias for articles in natalizumab set................................................ D-4
Appendix Table D3. KQ1 risk of bias for articles in pregnancy set.................................................. D-5
References for Appendix D ........................................................................................................... D-6
<table>
<thead>
<tr>
<th>Study Country Funding</th>
<th>Type of Study</th>
<th>Overall Risk of Bias Assessment</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Glatiramer Acetate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debouverie, 2007¹ France Not reported</td>
<td>Prospective observational</td>
<td>High</td>
<td>Unclear inclusion criteria (25% “contraindication to interferon beta” not defined), unreliable outcome measure (relapse per MD), numerous unblinded outcomes assessors, no assessment of impact of attrition</td>
</tr>
<tr>
<td>Ford, 2010² (Johnson) United States Industry</td>
<td>Prospective open-label</td>
<td>High</td>
<td>Attrition, unblinded outcomes assessment</td>
</tr>
<tr>
<td>Miller, 2008³ United States Industry</td>
<td>Followup of compassionate use</td>
<td>High</td>
<td>Attrition, no assessment of impact of attrition, small sample size, unequal length of followup, competing exposures not described</td>
</tr>
<tr>
<td><strong>Teriflunomide</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contafavreux, 2012² N America, Europe Industry</td>
<td>Open-label extension of RCT: randomized placebo group</td>
<td>High</td>
<td>Attrition, no assessment of impact of attrition, reporting bias (Table 2, Fig. 2 by extension-phase drug only)</td>
</tr>
<tr>
<td><strong>IFNβ1a</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uitdehaag, 2011⁴ (PRISMS, Gold) 22 centers: Europe, Canada, Australia Not funded</td>
<td>Retrospective long-term followup of open-label extension of mixed design</td>
<td>High</td>
<td>High attrition, no assessment of impact of attrition, competing exposures not described, potential for recall bias (relapse outcome), small sample, post-hoc analyses</td>
</tr>
<tr>
<td><strong>IFNβ1b</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bencsik, 2006⁵ Hungary Not reported</td>
<td>Prospective cohort</td>
<td>High</td>
<td>Attrition, small/insufficient sample size, unblinded outcomes assessment, no comparator</td>
</tr>
<tr>
<td>Carmona, 2008¹ Spain Not reported</td>
<td>Prospective cohort with historical control</td>
<td>High</td>
<td>Unclear selection of historical controls, insufficient reporting of baseline characteristics, attrition</td>
</tr>
<tr>
<td>Goodin, 2012⁶ Reder, 2010⁷ North America Industry</td>
<td>Long-term followup of RCT</td>
<td>Moderate</td>
<td>Competing exposures not described or included in the analysis</td>
</tr>
<tr>
<td>Kappos, 2009⁸ N America, Europe Industry</td>
<td>Open-label extension of RCT</td>
<td>High</td>
<td>Attrition, no assessment of impact of attrition</td>
</tr>
<tr>
<td>Rio, 2007¹¹ Spain Not reported</td>
<td>Post-marketing study</td>
<td>High</td>
<td>Attrition, baseline differences in case-mix between groups were not accounted for in the analysis.</td>
</tr>
<tr>
<td><strong>Interferon beta mixed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patti, 2006¹² Italy, 2 sites Reported not industry</td>
<td>Open-label observational</td>
<td>High</td>
<td>Attrition, no assessment of impact of attrition, no attempt to balance differences in baseline MS characteristics across groups, inconsistent reporting (text vs. Table 1), unblinded outcomes assessment</td>
</tr>
<tr>
<td>Study Country Funding</td>
<td>Type of Study</td>
<td>Overall Risk of Bias Assessment</td>
<td>Rationale</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------</td>
<td>---------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Portaccio, 2008&lt;sup&gt;13&lt;/sup&gt; Italy Not reported</td>
<td>Retrospective cohort</td>
<td>High</td>
<td>Unequal length of treatment, unequal length of followup, lack of detail about variables used in Cox models and attempts to deal with baseline differences in MS characteristics, questionable control of confounding factors in analysis</td>
</tr>
<tr>
<td>Shirani, 2012&lt;sup&gt;14&lt;/sup&gt; Canada Government</td>
<td>Retrospective cohort (database) with historical and contemporary control groups</td>
<td>Moderate</td>
<td>Lack of detail in describing the intervention (all IFNβs considered as one therapy, breaks in treatment ignored), competing exposures not described, unblinded assessors, sample selection (limits of database capture).</td>
</tr>
<tr>
<td><strong>Mixed Drugs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tedeholm, 2013&lt;sup&gt;15&lt;/sup&gt; Sweden Government</td>
<td>Retrospective cohort (database) with historical control group</td>
<td>High</td>
<td>Selection bias, lack of generalizability of contemporary treated group to population-based historical control, probable treatment by indication</td>
</tr>
<tr>
<td>Rio, 2005&lt;sup&gt;16&lt;/sup&gt; Spain Not reported</td>
<td>Open-label nonrandomized post-marketing observational</td>
<td>High</td>
<td>Baseline differences in case-mix between groups were not accounted for in the analysis, unequal length of treatment across groups, unequal length of followup, incomplete and unclear tracking of patients over time (table values conflict with text), competing exposures not described, attrition</td>
</tr>
<tr>
<td>Bergamaschi, 2012&lt;sup&gt;17&lt;/sup&gt; Italy Unfunded</td>
<td>Retrospective cohort (database) categorized by Bayesian modeling for risk factors</td>
<td>Unclear</td>
<td>Unclear if modeling accounted for treatment by indication. Nearly 30% of the remaining registry cohort was excluded because they had not been treated, or had not been treated with a first-line therapy, or had also been treated with a second-line therapy.</td>
</tr>
</tbody>
</table>
### Appendix Table D2. KQ1 risk of bias for articles in natalizumab set

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Funding</th>
<th>Type of Study</th>
<th>Overall Risk of Bias</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ARR Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jokubaitis, 2014a</td>
<td>Australia and international Government, non-profit, industry</td>
<td></td>
<td>Retrospective cohort</td>
<td>High</td>
<td>Retrospective, large sample with lack of individual-level outcome assessment, relapse severity measured by proxy (relapse treatment: ambulatory, hospitalization or none)</td>
</tr>
<tr>
<td>Havla, 2013b</td>
<td>Germany</td>
<td>Not reported</td>
<td>Retrospective cohort</td>
<td>High</td>
<td>Retrospective, inadequate sample size, outcome (rebound definition) not clearly specified, confounding variables not addressed (reason for interruption, prior treatment)</td>
</tr>
<tr>
<td>Havla, 2011c</td>
<td>Germany</td>
<td>Not reported</td>
<td>Prospective cohort</td>
<td>High</td>
<td>Inadequate sample size, outcome (rebound definition) not clearly specified, confounding variables not addressed (reason for interruption, prior treatment)</td>
</tr>
<tr>
<td>Kaufman, 2011d</td>
<td>United States</td>
<td>Not funded</td>
<td>Retrospective chart review</td>
<td>High</td>
<td>Retrospective, outcome (rebound definition) not clearly specified, confounding variables not addressed (reason for interruption, prior treatment)</td>
</tr>
<tr>
<td>Kerbrat, 2011e</td>
<td>France</td>
<td>Not reported</td>
<td>Retrospective chart review</td>
<td>High</td>
<td>Retrospective, outcome (rebound definition) not clearly specified, attrition in MRI outcomes, confounding variables not addressed (prior treatment)</td>
</tr>
<tr>
<td>O’Connor, 2011f</td>
<td>Canada</td>
<td>Industry</td>
<td>Prospective post hoc analysis of trial data</td>
<td>High</td>
<td>Attrition in MRI outcomes, large sample with lack of individual-level outcome assessment, confounding variables not addressed (prior treatment)</td>
</tr>
<tr>
<td>Rossi, 2013g</td>
<td>Italy</td>
<td>Industry</td>
<td>Prospective cohort</td>
<td>High</td>
<td>Outcome (rebound definition) not clearly specified, attrition in MRI outcomes, confounding variables not addressed (prior treatment)</td>
</tr>
<tr>
<td>Sorensen, 2014h</td>
<td>Denmark</td>
<td>Not reported</td>
<td>Retrospective; national registry</td>
<td>High</td>
<td>Retrospective, confounding variables not addressed (reason for interruption, prior treatment)</td>
</tr>
<tr>
<td><strong>MRI Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boriello, 2012i</td>
<td>Spain</td>
<td>Not reported</td>
<td>Prospective cohort post-marketing</td>
<td>High</td>
<td>Inadequate sample size, outcome (rebound definition) not clearly specified, confounding variables not addressed (baseline EDSS 5.0, prior treatment) inadequate reporting of relapses</td>
</tr>
<tr>
<td>Boriello, 2011j</td>
<td>Spain</td>
<td>Not funded</td>
<td>Prospective cohort post-marketing</td>
<td>High</td>
<td>Inadequate sample size, outcome (rebound definition) not clearly specified, unequal length of MRI followup, confounding variables not addressed (prior treatment)</td>
</tr>
<tr>
<td>Miravalle, 2011k</td>
<td>United States</td>
<td>Industry</td>
<td>Prospective cohort</td>
<td>High</td>
<td>Inadequate sample size, outcome (rebound definition) not clearly specified, inadequate reporting of relapses, confounding variables not addressed (prior treatment)</td>
</tr>
<tr>
<td>Rossi, 2013l</td>
<td>Italy</td>
<td>Industry</td>
<td>Prospective cohort</td>
<td>High</td>
<td>Outcome (rebound definition) not clearly specified, attrition in MRI outcomes, confounding variables not addressed (prior treatment)</td>
</tr>
</tbody>
</table>

ARR=annualized relapse rate; EDSS=Expanded Disability Status Scale; MRI=magnetic resonance imaging
### Table Appendix D3. KQ1 risk of bias for articles in pregnancy set

<table>
<thead>
<tr>
<th>Study Country Funding</th>
<th>Type of Study</th>
<th>Overall Risk of Bias Assessment</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interferon beta 1a</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandberg-Wollheim, 2011&lt;sup&gt;29&lt;/sup&gt; Sweden Industry</td>
<td>Prospective global drug safety database</td>
<td>High</td>
<td>Unintended exposures not described</td>
</tr>
<tr>
<td>Sandberg-Wollheim, 2005&lt;sup&gt;30&lt;/sup&gt; Sweden Industry</td>
<td>Prospective individual patient data from 8 trials</td>
<td>High</td>
<td>Insufficient sample size, unintended exposures not described</td>
</tr>
<tr>
<td><strong>Interferon beta mixed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amato, 2010&lt;sup&gt;31&lt;/sup&gt; Italy Not reported</td>
<td>Prospective cohort, 21 clinics</td>
<td>High</td>
<td>Insufficient sample size, lack of detail in describing the intervention (e.g., specific drug) in results, lack of detail in describing statistical methods</td>
</tr>
<tr>
<td>Boskovic, 2005&lt;sup&gt;32&lt;/sup&gt; Canada Not reported</td>
<td>Prospective cohort, women who contacted risk counseling program</td>
<td>High</td>
<td>Insufficient sample size, lack of detail in describing the intervention (e.g., specific drug), inappropriate comparison, statistical methods to assess the outcome used inappropriately, lack of interpretability</td>
</tr>
<tr>
<td>Patti, 2008&lt;sup&gt;33&lt;/sup&gt; Italy Not reported</td>
<td>Retrospective cohort, one clinic</td>
<td>High</td>
<td>Retrospective design, insufficient sample size, lack of detail in describing the intervention (e.g., specific drug), lack of interpretability</td>
</tr>
<tr>
<td><strong>Interferon beta mixed, glatiramer acetate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finkelsztejn, 2011&lt;sup&gt;34&lt;/sup&gt; Brazil Not reported</td>
<td>Retrospective, voluntary database</td>
<td>High</td>
<td>Insufficient sample size, lack of detail in describing the intervention (e.g., specific drug), unintended exposures (e.g., smoking) not described, lack of interpretability</td>
</tr>
<tr>
<td>Fragoso, 2013&lt;sup&gt;35&lt;/sup&gt; International</td>
<td>Retrospective, voluntary database</td>
<td>High</td>
<td>Selectively recruited participants, insufficient sample size, lack of detail in describing the intervention (e.g., specific drug), unintended exposures (e.g., smoking) not described, unclear length of followup for maternal outcomes</td>
</tr>
<tr>
<td>Hellwig, 2012&lt;sup&gt;36&lt;/sup&gt; Germany Not funded</td>
<td>Retrospective, risk counseling database</td>
<td>High</td>
<td>Insufficient sample size, lack of detail in describing the intervention (e.g., specific drug), unintended exposures (e.g., smoking) not described</td>
</tr>
<tr>
<td>Lu, 2012&lt;sup&gt;37&lt;/sup&gt; Canada Government</td>
<td>Retrospective cohort, provincial database</td>
<td>High</td>
<td>Insufficient sample size, lack of detail in describing the intervention (e.g., specific drug)</td>
</tr>
<tr>
<td>Weber-Schoendorfer, 2009&lt;sup&gt;38&lt;/sup&gt; Germany</td>
<td>Prospective cohort, risk assessment program</td>
<td>High</td>
<td>Insufficient sample size, lack of detail in describing the intervention (e.g., specific drug), inappropriate comparison</td>
</tr>
<tr>
<td><strong>Glatiramer acetate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giannini, 2012&lt;sup&gt;39&lt;/sup&gt; Italy Not reported</td>
<td>Prospective cohort, 21 clinics</td>
<td>High</td>
<td>Insufficient sample size</td>
</tr>
<tr>
<td><strong>Natalizumab</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hellwig, 2011&lt;sup&gt;40&lt;/sup&gt; Germany Not funded</td>
<td>Prospective counseling database</td>
<td>High</td>
<td>Insufficient sample size, unintended exposures (e.g., smoking) not described</td>
</tr>
</tbody>
</table>
References for Appendix D


