A prototype Web platform to facilitate public engagement with medical evidence about rheumatoid arthritis medications

by
Adam Baker

A thesis
presented to the Independent Studies Program
of the University of Waterloo
in fulfillment of the
thesis requirements for the degree
Bachelor of Independent Studies (BIS)

Waterloo, Ontario, Canada
2016

Author's declaration

I hereby declare that I am the sole author of this research paper.

I authorize the University of Waterloo to lend this research paper to other institutions or individuals for the purpose of scholarly research.

April 14, 2016

I further authorize the University of Waterloo to reproduce this research paper by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

April 14, 2016

Abstract

Contemporary technologies and user interface design enable people to routinely interact with data in their everyday lives. While consumer applications for shopping and travel often feature data-driven user interfaces, health resources rarely do. These resources rely on manual translation of medical evidence into prose instead of providing users the capacity to interact with underlying data. The abstraction away from details about treatment options, including data about efficacy, harms, and patient-reported outcomes, stands in the way of people who may wish to become fully informed when taking on important medical decisions. In spite of barriers that restrict access to and potential to apply medical evidence, this project explored whether contemporary open-source Web technologies could be adapted to create data-driven resources for the exploration of such evidence.

A prototype platform and example applications were developed using JavaScript and React.js, with Google Spreadsheets as a data store for medical evidence related about twelve disease-modifying antirheumatic drugs (DMARDs) commonly used to treat rheumatoid arthritis. Research findings were manually encoded from diverse sources, and a controlled vocabulary and data visualization components built to bridge the gap between outcomes and data publishing formats favored in research, and issues important to patients with rheumatoid arthritis. The volume and heterogeneity of source evidence revealed no straightforward parallel to consumer data-driven online applications, especially where evidence conflicts or is uncertain. Nevertheless, this thesis demonstrates that extant and ready-made technologies can be combined to create an extensible, data-driven platform and user interface elements to investigate and visualize certain kinds of evidence about chronic disease treatment options. Future research might investigate how such platforms might be incorporated into patient-facing decision aids, automated synthesis of research findings, and collaborative tools to encode evidence.

Acknowledgements

This project would not have been possible without the support of a great number of people. Countless thanks are owed to my supervisors Dr. Thomas Agoritsas at McMaster University, Dr. Jennifer Barton at Oregon Health and Science University, and Dr. Jim Wallace at the University of Waterloo. Each was patient as the project was extended and its scope changed—and as I increased the length of this paper. They provided invaluable advice in domains as diverse as usability, data visualization, evidence-based medicine, decision aids, and clinical experience of treating rheumatoid arthritis. Their feedback and commitment to me, from proposal to thesis, was humbling. Thank you to Dr. Victor Montori of the Mayo Clinic for introducing me to Dr. Agoritsas and Dr. Barton, and to Dr. Mark Havitz of the Academic Board of Independent Studies for introducing me to Dr. Wallace, and for serving as my Academic Board advisor.

The Academic Board of Independent Studies was both encouraging and helpfully critical at the outset of my thesis project. They foresaw my project's enormous scope before I did. The Independent Studies program at the University of Waterloo offered me an opportunity to forge my own unique path in design. Thank you to Susan Gow who guided me on the the program's path towards this thesis project. My first Independent Studies advisor Mike Elmitt, Professor Emeritus at the University of Waterloo School of Architecture, expanded my sense of design in every way, and my most recent advisor, Dr. Ted McGee of Independent Studies, inspired confidence in my capacity to complete this thesis.

Many mentors, friends, and family have supported me in this process—technically and otherwise—including Tony Dang, Thomas Goetz, Dr. Taha Kass-Hout, Akos Kokai, Skye Louis, Evan Martin, Matt Mohebbi, Sarah Nahm, Nadav Savio, Eric So, Dr. Meena Tappouni, and Chris Young. A very special thanks to my first proofreader, my mother Judy Cipin.

Finally, I am grateful to all the people who have labored to help everyone navigate medical decisions.

Table of contents

List of figures

List of tables

List of frequently used abbreviations

ACR. *American College of Rheumatology*.
AE. *Adverse event or adverse effect*.
API. *Application programming interface*.
DA. *Decision aid*.
bDMARD. *Biologic disease-modifying antirheumatic drug*.
csDMARD. *Conventional synthetic disease-modifying antirheumatic drug*.
DMARD. *Disease-modifying antirheumatic drug*.
EBM. *Evidence-based medicine*.
FDA. *Food and Drug Administration*.
RA. *Rheumatoid arthritis*.
SDM. *Shared decision making*.
SPL. *Structured Product Labeling*.
UI. *User interface*.
UX. *User experience*.
VAS. *Visual analog scale*.

Nomenclature

Who is a "patient?"

The term "patient" is often used to describe someone involved in an episode of care (Greene & Partridge, 2007) and seen to be in a "sick" role. In general, I do not believe that people living with ordinary health issues—as most of us do—should be referred to as *patients* in the context of medical information, education, and resources. However, it is awkward to find the right noun that isn't the word *patient* in some cases. Through this document, *patient, person, non-specialist, member of the public*, and similar plural terms are used and refer to the same notion—a person living with health information needs. We are all a patient at some point, but hardly every day. Occasionally, the terms *caregiver* or *family member* are mentioned to invoke the important notion that many people do not make medical decisions alone, but in concert with loved ones or other people in their lives. Similarly, the terms *practitioner, clinician, and provider* are often used in place of *physician, doctor, nurse, pharmacist*, and so on, in order to avoid awkward linguistic gymnastics and to recognize situations in which many care providers work together on behalf of patients.

Non-gendered pronouns

To respect gender preferences and avoid complicated use of *he* and *she,* the non-gendered singular pronoun *they* is substituted when a gender is not implied. For example, "if a rheumatologist does not recommend a specific medication, they may offer their patient a decision aid."

Numerals

Numbers are frequently cited in this paper. Although it may be customary to write out numbers below 10, such writing might be confusing or unclear—for example, *nausea was experienced by 11 percent of patients, while sleep problems affected three percent of men and 14 percent of women.* To ensure clarity, numerals are used instead of writing numbers out.

Chapter 1

Introduction

- The major work of this thesis project is a prototype platform to support public engagement with medical evidence. An example application—called the *Navigator* prototype—can be accessed at:
  http://thesis.merges.net/navigator
- Data are encoded in a publicly visible Google Spreadsheet at:
  http://goo.gl/gdbVuf
- The prototype is open source. Its code is checked into GitHub at:
  http://github.com/merges/abist

The rise in potential medical information needs

People increasingly face complex medical treatment decisions related to chronic illness. Chronic conditions—arthritis, asthma, cancer, chronic obstructive pulmonary disease (COPD), coronary heart disease, diabetes, hepatitis, hypertension, and chronic kidney disease—are widely prevalent and often comorbid, meaning that they frequently accompany one another. About half of American adults lives with one of those chronic health conditions, and 1 in 4 with more than one (Ward, Schiller, & Goodman, 2014), to say nothing of other chronic health issues like depression, pain, and anxiety. Medical research has advanced new and sometimes superior treatment options for many of these conditions, slowing or stopping their deleterious effects and permitting extended and higher quality of life. The obvious benefit of this medical progress is that not only can people live with many of these conditions, but live better and longer. However effective some of these interventions may be, they often come along with adverse events, and may be difficult for someone to integrate into their life (Leppin, Montori, & Gionfriddo, 2015).

Alongside the rise of treatment advances for these chronic conditions, medical decision making has changed. An approach called *evidence-based medicine* (EBM) has emerged and contributed to the mushrooming of medical research data. It asserts that medical decision making based on the best available data from high-quality clinical trials and epidemiological research about what works best, instead of strictly clinical experience or pathophysiologic rationale (that is, theories about the mechanics of illness and treatment) is preferable (Guyatt, Cairns, & Churchill, 1992).

Formerly paternalistic "doctor-knows-best" medicine has somewhat given way to more deferential and balanced models. In the widely adopted *informed consent* model, practitioners retain decision authority, but tell patients about the nature, benefits, and potential harms of medical interventions—tests and treatments—and obtain permission to proceed from these "informed" patients (Emanuel & Emanuel, 1992). Recently, the notion of *shared decision making* (SDM) has gained momentum. Advocates urge that practitioners, patients, and family members or caregivers collaborate and make medical decisions *together*, rather than vesting authority solely in the practitioner or the patient (Charles, Gafni, & Whelan, 1997; Stiggelbout et al., 2012). Some suggest that the same evidence that underpins EBM practice is an essential part of well-informed SDM (Barratt, 2008; Epstein & Gramling, 2013). However, the practitioner almost always has better access to such information than their patient, an example of information asymmetry. To make a fully informed decision, the patient and their clinician would ideally share access to the same best available evidence to answer questions and explore options.

People might expect that with all the research about new treatments, medicine as an institution might now know all about which interventions work best in all situations. They might expect that there are clear "winning" treatments, and clear answers to almost any medical question. In some cases, that is true—and fortunate. In many cases, the picture is cloudy (Epstein & Gramling, 2013). A new treatment might be slightly more effective in one way than an established treatment, but carry different risks and cost more. Or all the treatment options for a condition might be substantially the same, where none of them work best. Or there may be no evidence about the impact of one course or another on outcomes that truly matter to the patient. Or the evidence may not be reliable because of bias, sourced from studies with non-representative patients, insufficient sample sizes, surrogate endpoints, or for a number of other reasons (Greenhalgh, 2014; Gordon Guyatt, Rennie, Meade, & Cook, 2008; Pannucci & Wilkins, 2010). These cases introduce *clinical equipoise*, a situation in which no single option is clearly the best choice, either because there is insufficient evidence to support one course over another, or because even if one *is* more effective for a desirable positive outcome, that its potential risks or harms may outweigh the benefit given the health status and preferences of a particular patient. Uncertainty like this plays a major role in medical decisions (G. Elwyn, Edwards, Kinnersley, & Grol, 2000), especially when such uncertainty prevails and people may be substantially under-informed about their options.

When evidence *is* available, it rarely exists in a form accessible and useful to patients. Researchers produce findings on outcomes that advance research interests. Those may align with concerns real people have, but they may not. Surrogate outcomes or endpoints—measurable but indirect indicators of treatment effectiveness which may not represent real patient interests—are common (Greenhalgh, 2014; Guyatt et al., 2008). Effective translation

is necessary to make evidence understandable to a wide audience, so that it can furnish support to people making medical decisions—from the jargon of medical research to the plain language and issues that matter to patients. Furthermore, most details of findings (e.g. statistics and estimates of effect) are locked in the prose of peer-reviewed articles in medical journals. Sometimes findings are pooled and synthesized into useful data-driven systematic reviews or meta-analyses. Even better—but unfortunately rarely—they may be encoded as machine-readable data that power interactive software for practitioners and their patients. In machine-readable form, evidence can be computed, translated, and visualized in many ways. Tools built with such data can tailor presentation of evidence to the needs of the individual seeking information—– and kept up to date as new evidence is produced. As that is an uncommon scenario, much evidence remains indecipherable for patients as medical jargon, in expensive access-controlled academic journals. Only bespoke, expert-crafted summaries for lay audiences, with findings frozen from the source evidence contemporaneous with their production, typically find general circulation among patients. People increasingly search online for medical information, including about treatment options (Fox & Duggan, 2013). When they search Google to equip themselves to become informed about their experience and what may be best for them, they encounter the implications of this outmoded way of publishing findings—generally *either* data-impoverished resources (conventional online resources like WebMD) in the case of the most common general searches, or the scholarly literature for less common search queries.

In non-medical domains, every day non-experts in the general public realize meaningful information from vast quantities of technical data—through the power of carefully designed user interfaces (UI). Using these data-driven UIs, they can employ applications in decision support. In a smartphone map app, huge volumes of data—mapping, historical and real-time traffic, transit route, schedule and real-time transit departure, among other types—are brought together and "translated" by a well-designed UI to provide a straightforward user experience (UX) for navigation by public transit. Web-based flight search tools combine, translate, and facilitate manipulation of similarly extensive technical data, such that people can find the most appropriate flight for their needs. In both of these cases, the source (sometimes called *raw*) data would themselves be difficult to use or useless without skillfully designed UI. The raw data must first be made machine-readable. Then each source can be individually presented in an appropriate form—map, table, or otherwise. Going further, they can be filtered and commingled and *re*-presented in new visual and audible ways, and made interactive, responsive to individual queries. This approach is customary in many domains of online information-seeking. Or, the data could be manually interpreted, written into comprehensive narrative prose and published, foregoing the inclusion of real-time updates—the prevailing approach in online health information, including public-facing information about medication effects.

Medical information is an online outlier: When people search to learn more about medications, they rarely find data-driven tools. Instead, they find manually translated articles—published only if certain entities (whether *WebMD, The Mayo Clinic,* or a health agency) have invested in the human effort to interpret evidence and synthesize it into a static, non-interactive resource. In this way, medical evidence is routinely translated for non-expert consumption. Work in other domains have demonstrated that complex technical data can be "brought to life" in interactive tools. These two concepts must be brought

together, to make data-driven, up-to-date interactive tools that provide similar value for medical information-seeking. That way practitioners and their patients can share a similar evidence base, partially addressing the in-clinic information asymmetry and facilitating truly informed SDM. Evidence-based shared decision making could increase self-efficacy, belief in treatment, goal concordance, and build stronger therapeutic alliances. It may also have detrimental effects, if patients discover data that raises confusion or overwhelms them.

It is not enough to imagine and describe such tools. They must be demonstrated, applying contemporary design and technologies to tackle basic problems, experiment with solutions, and prove out concepts. In fact, since there are many parties involved in the production and consumption of the medical knowledge that would underlie such tools—from researchers to patients, from doctors to programmers, from designers to caregivers—many dimensions of the problem must be explored simultaneously, so that each party can understand the value and possibility. The central problems include:

- Can pertinent medical evidence be encoded and made machine readable?
- Can this evidence be built into a system that can "translate" it so that it is relevant to medical decision making?
- Can contemporary design principles be applied to make the data or evidence understandable?

Ultimately, it is up to individuals to decide whether or not they want this kind of information; not every person desires or requires it. Some patients are highly motivated, and today jump through many hoops and surmount literacy challenges to find and interpret medical evidence. Others will continue to rely on their doctor's educated and trustworthy knowledge, and some on plain language summaries they find in the most popular online search results. But there is no reason that efforts should not be made to learn about how to develop digital tools for people who *do* want evidence-based, tailored information to help them understand their choices and make decisions—especially when those decisions are complex and the choices unfamiliar. One such diagnosis after which people may face these kinds of decisions is rheumatoid arthritis (RA).

Rheumatoid arthritis and complex medical decisions

The 13th edition of the Primer on the Rheumatic Diseases (Klippel, Stone, Crofford, & White, 2008) supplies a comprehensive background on the clinical characteristics, pathophysiology, and epidemiology of rheumatoid arthritis in several book sections (Oliver & St Clair, 2008; Tehlirian & Bathon, 2008; Waldburger & Firestein, 2008). RA is a chronic autoimmune condition of unknown cause. It is characterized by an inflammatory process in the synovium, the tissue between the joint capsule and joint cavity that permits smooth and full range of motion in synovial joints—the most common and movable in the body, including those of the hands, wrists, shoulders, elbows, hips, knees, ankles, and feet—but also involves other organs through systemic inflammation. As RA manifests over weeks or months, joints become swollen, painful, and stiff, and is accompanied by general fatigue. Morning stiffness—often for up to 2 hours—is common. These symptoms impair a person's ability to perform ordinary activities, from cooking to bathing to exercising to working. People may initially seek treatment for their symptoms—to relieve pain and swelling—but if the underlying RA disease process is untreated, it may destroy not only the synovium, but soft joint tissue (cartilage, ligaments, and tendons) and even bone. Such destruction results in permanent painful, disabling, and disfiguring joint damage. RA is a chronic condition, though it does not necessarily follow a single course. It is generally progressive, with periods of quiet and flares of disease activity.

Helmick et al. (2008) estimate that RA affects about 0.6% of the U.S. adult population, or approximately 1.3 million people. Women are disproportionately affected, at a rate around two to three times that of men (Klippel et al., 2008). Genetic predispositions to RA are almost certainly involved, though it remains uncertain what "sets off" the autoinflammatory process. Increasing age is a risk factor, as is smoking, the only associated environmental risk factor (Klippel et al., 2008). The disease process may be active for a number of weeks or months before a patient is referred to a rheumatologist, by which time the signs and symptoms may clearly point to RA. Laboratory tests for biomarkers like rheumatoid factor (RF) or anti-cyclic citrullinated peptide (anti-CCP) may also be performed to add to the evidence for an RA diagnosis (Klippel et al., 2008).

Treatment

Because RA is a chronic disease of systemic inflammation that goes beyond the joints, it warrants holistic care involving more than just medicines. Symptom control—of pain, swelling, and fatigue—is important especially early in treatment or during a flare up, but so may be limiting stress, adjusting diet, exercising, managing weight, modifying work and leisure activities, building an emotional support network, and more. Crucially, the underlying disease process must be treated in order to slow or prevent joint damage and other sequelae (medical conditions that are a consequence of RA) (Klippel et al., 2008). The disease process can be slowed or arrested by one or more of a host of medicines, called disease-modifying antirheumatic drugs (DMARDs). Many of the medications that relieve acute symptoms are not DMARDs—for example,

over-the-counter anti-inflammatory drugs. See Table 1 for a sample of some of the most common drugs used to treat RA.

In the past several decades, medical therapy for RA has changed substantially. New treatments have emerged approximately every quarter century for the past hundred years: Aspirin at the turn of the 20th century, gold salts in the late 1920s, corticosteroids and sulfasalazine in the 1950s, use of methotrexate in the 1980s, and biologic DMARDs in the 2000s (Abramson, n.d.). Thanks to a better understanding of the disease process, strategies for its treatment, and in some cases new medications that target specific pro-inflammatory factors in RA, it is possible to significantly slow disease progression (Klippel et al., 2008; Smolen et al., 2013)—especially important for people with RA since it is associated with significant premature mortality from multiple complications, including cardiovascular disease (Kvien, 2004). However, for many people treatment is a winding road. The first treatment may not work well, or it may not work on a long-term basis. DMARDs are most certainly not free of harms, ranging from nausea to liver injury to serious infection resulting from immunosuppressant action. These adverse reactions may make a drug intolerable or harmful.

Table 1
Medications commonly used to treat rheumatoid arthritis.

Kind	Purpose	Subtype	Route of administration	Examples	Common risks and side effects
Disease-modifying antirheumatic drug (DMARD)	Slow disease progression, consequently relieving symptoms and preventing long-term damage.	conventional synthetic DMARD (csDMARD)	Usually oral, sometimes injection	Auranofin (Ridaura) Azathioprine (Imuran) Hydroxychloroquine (Plaquenil) Leflunomide (Arava) Methotrexate (Rheumatrex) Sulfasalazine (Azulfidine)	Immune suppresion, gastrointestinal issues, liver injury (methotrexate, leflunomide)
		Biologic agent (bDMARD)	Injection or infusion	Abatacept (Orencia) Adalimumab (Humira) Anakinra (Kineret) Certolizumab (Cimzia) Etanercept (Enbrel) Golimumab (Simponi) Infliximab (Remicade) Rituximab (Rituxan) Tocilizumab (Actemra)	Immune suppression, serious infection
		Janus kinase (JAK) inhibitor	Oral	Tofacitinib (Xeljanz)	
Corticosteroid	Relieve acute severe inflammation, especially RA "flare-ups."		Oral or injection	Cortisone Dexamethasone Hydrocortisone Prednisone	Immune suppression, loss of bone density, weight gain
Non-steroidal anti-inflammatory drug (NSAID)	Relieve symptomatic inflammation and pain.		Oral	Aspirin Celecoxib (Celebrex) Diclofenac (Zorvolex) Ibuprofen (Motrin, Advil) Indomethacin (Indocin) Naproxen (Aleve, Naprosyn)	Bleeding, gastrointenstinal issues
Analgesic	Relieve acute pain.	Non-opioid	Oral	Acetaminophen (Tylenol)	Liver injury
		Opioid (and opioid combination)	Oral	Acetaminophen/Codeine (Tylenol #3) Aspirin/Oxycodone (Percodan) Fentanyl Hydrocodone Hydromorphone (Dilaudid) Morphine (MS Contin) Ibuprofen/Hydrocodone (Vicoprofen) Oxycodone (Ocycontin) Tramadol (Ultram)	Physical dependence and abuse, gastrointestinal issues

Treatment frequently involves polypharmacy (the use of multiple medications at once), raising the risk of interactions and hazards related to drug interactions or inadvertent incorrect use. Use of five or more medications at once is common, although some may be taken for comorbid conditions rather than for RA (Filkova et al., 2015; Treharne et al., 2007). Longer-term use (months to years) of DMARDs may be combined with short-term use (weeks to months) of corticosteroids to quickly relieve inflammation or augment the primary DMARD, and over-the-counter (OTC) or more potent prescription medications to relieve pain (Klippel et al., 2008). The advantage of recent medical progress is that if the medications work well, some people can experience remission and defer longer-term health consequences. The disadvantage is that the treatment process can be substantially more complicated to manage, with significant impacts on daily life—all of which must be considered in the context of living with a condition that itself has a huge impact on someone's life (Barton, 2009). The challenge of recruiting cooperation into an optimal care plan is complex. As Matteson (2001) says, "The health-related beliefs, goals, and desires of patients are important predictors of compliance with treatment and outcome. A willingness on the part of health care providers to understand and work with these beliefs, and to educate patients about the disease, is as fundamental to the successful treatment of RA as any medication that can be prescribed." Recent clinical practice guidelines entreat physicians to engage in such shared decision making, with the European Union League Against Rheumatism treatment recommendations saying that "treatment of RA patients should aim at the best care and must be based on a shared decision between the patient and the rheumatologist," and going as far as to note that "the Task Force [who created the guidelines] decided that decision-sharing by patient and rheumatologist is of such overwhelming importance that it should spearhead the recommendations" (Smolen et al., 2013).

Disease-modifying antirheumatic drugs (DMARDs)

DMARDS are agents that act on the disease process underlying RA. They can slow disease progression, reduce inflammation, and prevent or limit joint damage. Consequently, they can lead to RA symptom relief. Generally, they work by suppressing or inhibiting various immune functions that would otherwise be involved in the inflammatory and destructive RA disease process, through a range of mechanisms of action (Brenner & Stevens, 2013). DMARDs vary widely in terms of side effect profile, route of administration, contraindications, cost, and impact on daily routine. Two commonly used types of DMARD are conventional synthetic DMARDs (csDMARDs) and biologic agents (bDMARDs). Most conventional DMARDs are taken orally, and are available in generic versions. Biologic DMARDs—so-called because they are engineered using living tissues and mimic substances like human antibodies—are a newer class of medication, first approved for use in the late 1990s. Biologic agents target a specific mechanism in the immune response in RA. Most are protected by patent (some expiring within months of this research paper's writing) and therefore only available as relatively costly brand name products. The first generic (called *biosimilar*) versions are just being approved and may alter the treatment landscape by increasing access to such DMARDs. csDMARDs are the mainstays of treatment, though biologics play an important role, especially when first-line drugs like methotrexate do not work well (Klippel et al., 2008; Singh et al., 2016; Smolen et al., 2013). DMARDs can produce adverse effects that range from the relatively benign (nausea and

gastrointestinal issues) to very serious (liver damage, immune suppression, and rarely, development of certain cancers).

Glucocorticoids (corticosteroids)

Glucocorticoids also have an effect on the underlying RA disease process. These drugs have wide systemic effects, including inhibition of proinflammatory factors and lowering levels of chemicals that are involved in immune response. While they can quickly and significantly reduce inflammation and provide relief, at larger doses they can cause serious harms (Brenner & Stevens, 2013), so they are typically used alongside DMARDs either early in treatment to ease symptoms or to help control flare-ups of RA. They may also be used for a longer duration at low doses in concert with DMARDs, although there is debate over whether they are safe even at low doses (Smolen et al., 2013).

Non-steroidal anti-inflammatory drugs (NSAIDs)

This is a large group of drugs that includes OTC medications like aspirin, ibuprofen (Advil or Motrin), and naproxen (Aleve), and prescription medications such as celecoxib (Celebrex). They can reduce inflammation and pain associated with RA, but do not slow disease progression and joint damage, and are not considered DMARDs (Brenner & Stevens, 2013). Someone who finds relief from NSAIDs of early and mild symptoms could believe that they are better, or getting better, but an unrecognized disease process could still be progressing toward joint damage.

Medical treatment strategies with DMARDs and corticosteroids

The right medical treatment depends on the interplay between the specific characteristics of someone's RA, their beliefs about the disease and treatment, and their life. Although there is no single best course of treatment, there are standards of care and up to date evidence-based clinical practice guidelines for RA treatment. They set out strategies and entire courses of treatment, with rules for evaluating, adding, removing, and switching medications. People with relatively mild RA might be treated with hydroxychloroquine or sulfasalazine, oral DMARDs with relatively few side effects. However, methotrexate is the "gold standard" and cornerstone of treatment for most people with RA. If someone doesn't respond to a low dose of methotrexate, its dose may be increased, or it may be supplemented with a low dose of a glucocorticoid like prednisone. For severe RA or if methotrexate does not work well, a biologic agent might be tried. This is a "step-up" approach to treatment, where DMARDs are added to a treatment regimen until the desired response is observed. A "step-down" approach would be to try more aggressive treatment to achieve an early response, and then lower doses or remove medications and still maintain the response. So-called "double therapy" and "triple therapy" using multiple oral DMARDs—methotrexate, sulfasalazine, and hydroxychloroquine—may confer a rapid benefit for aggressive RA. However, initial DMARD monotherapy remains the standard of care (Oliver & St Clair, 2008; Singh et al., 2016; Smolen et al., 2013).

A treat-to-target (TTT) strategy—similar to one that may be used when people are prescribed medications to reduce their blood pressure to a specific figure (e.g. below 140/90)—suggests that treatment should aim to achieve a specific,

measurable goal (Atar, Birkeland, & Uhlig, 2010). Rather than relieving symptoms or slowing disease progression *generally*, specific target values of certain measures or biomarkers are the goalposts. TTT depends on goals that can be derived from an extensive and high-quality (low bias and rigorous) evidence base that clearly shows the relationship between these markers and beneficial outcomes. The "ultimate goal" of RA treatment may be remission (Oliver & St Clair, 2008), but remission itself is difficult to define. In order to treat to target, surrogate indicators of low disease activity must be the targets. And while it may be that successful retardation of RA requires aggressive treatment to a target, that might involve polypharmacy with relatively toxic DMARDs that are difficult for some people to tolerate. The "right" targets would ideally be defined in a shared decision making process between clinician and patient, taking into account the patient's own goals and the potential for tolerable and appropriate treatments to help meet targets and the patient's needs. With surrogate endpoints—not necessarily patient-important or patient-defined outcomes—TTT and patient goals may not align well.

Regardless of the treatment approach, if someone does not take their medication regularly, as prescribed, the potential benefits can not be conferred. Beliefs about the medication—which may be influenced by a constellation of factors ranging from individual values to physician trust to medication knowledge—influence adherence to treatments. Horne & Weinman (1998) found that people who felt strongly that their treatment was necessary reported higher adherence rates, while those who has strong concerns about their medications (e.g. about side effects) reported lower adherence rates. Data-driven online resources might play a role in educating people about why medications are necessary to control RA, and how well they work. But they might also raise alarm and greater concerns. In principle, though, it is hopeful that an agreeable decision and knowledge about a treatment's benefits and harms could lead to greater confidence.

Preferences for treatment and involvement in treatment decisions

Patients do not have uniform needs and preferences for both treatments and involvement in decision making. In addition to concerns unique to each individual, there are documented differences among certain populations. One study demonstrated that black RA patients in the U.S., for example, prefer more conservative treatment and are more harm-averse than white patients, who are relatively more concerned with the potential benefits of treatment (Constantinescu, Goucher, Weinstein, Smith, & Fraenkel, 2009). Constantinescu et al. remark "studies attempting to explain racial variability in patient preferences have found that differences in spirituality, health beliefs, perceptions of benefit, and trust all influence patients' treatment preferences for medical interventions." In a review of patient preferences in RA treatment, Barton (2009) notes that variation in preferences and beliefs about medications may partially explain disparities in health outcomes, due to their influence on adherence. Socioeconomic factors—including educational attainment and household income—may account for significant disparities in RA outcomes in some populations (Baldassari et al., 2014) and hamper access to information and care, influencing preferences. Data-driven educational tools could play a role in "changing minds" about benefits (related to treatment necessity) and risks (related to treatment concerns) of RA treatments (Liana Fraenkel et al.,

2012), though it is very hypothetical to say that they might also play a role in reducing these outcome disparities.

Nota et al. (2014) found that around 60% of inflammatory arthritis patients prefer to participate in shared decision making, but point out that "Upon receiving the diagnosis, the patient needs to process a lot of information (about the influence of this chronic disease on daily life, starting aggressive treatment, etc.) in a short time. Not being aware of having a choice, little time, and/or an overload of information may be a barrier for patient involvement." For these people, resources they can process in a place and at a pace of their choosing may be beneficial—especially, perhaps, if they were shared by their physician. Perhaps the "short time" for processing could be extended.

Living with RA

This project addresses but a narrow "slice" of living with RA—seeking information about a new medical treatment. (And even then, only a subset of DMARDs.) The "best" care, and living with RA, is much larger than this decision. Hopefully, networks of friends, family, and healthcare professionals provide critical emotional support. Someone with RA may undertake significant lifestyle changes—some to accommodate symptoms of RA. Stress, diet, day-to-day symptoms, exercise, daily activities, and plenty of other more urgent considerations are likely to prevail over a few questions about medications. Nonetheless, because of the critical role that RA medications play in that larger milieu, such questions deserve attention by designers and researchers who make online resources for patients to learn about their options. Effective medical treatment is part of enabling someone with RA to thrive, and brings on real implications—such as a potential financial burden, regular laboratory tests to monitor for potential adverse effects (e.g. liver injury), and lifestyle changes (e.g. perhaps stopping drinking alcohol), to say nothing of unwelcome side effects and anxieties. The following scenarios illustrate complex medical decisions around starting and changing DMARD therapy.

Scenarios

Kim (new RA)

Kim, a married 36 year-old mother of two young children—a pre-schooler and a toddler—and professional caterer, has just received a diagnosis of RA from her rheumatologist, after several months of otherwise unexplained joint stiffness and fatigue that have interfered with Kim's daily activities. Since a visit to her primary care physician, she has been treating her pain and inflammation with over-the-counter ibuprofen. That hasn't been enough to help her pain and stiffness some days, and Kim has had to cancel several catering engagements

with clients to stay at home—foregoing her income and making her feel depressed. Her rheumatologist noted that Kim has a relatively high number of tender and swollen joints and severe self-reported pain and stiffness, which may be signs of early active disease. As such, the rheumatologist's recommendation was that she undergo treatment with methotrexate, taken orally once a week, in order to mitigate the disease process and hopefully bring Kim to a state of remission, in which she would be relatively free of symptoms and able to live comfortably. However, Kim and her husband have been planning to conceive, and methotrexate is not safe during pregnancy, nor lactation. Furthermore, Kim is not a fan of taking medication at all. She would rather treat any illness in a more holistic way, limiting use of medications. Methotrexate scares her, because she knows of it as a cancer treatment and has also heard of people injecting it. The rheumatologist explained that initial treatment with methotrexate is understood to be a relatively safe and effective way of achieving remission, though not benign—methotrexate carries with it the risk of side effects that may equally interfere with everyday life. It might take several months to take effect, and impact Kim's family planning. As she's in her mid-30s, Kim and her husband are wary of delaying their planned pregnancy.

Kim, her husband, and her rheumatologist are together facing a complex medical decision with multiple potential courses of action. They include:

- Treating Kim's RA now with methotrexate, and delaying their planned pregnancy. The benefits of this course of action may include rapid induction of disease remission and relief from pain, fatigue, and stiffness. But it is a course of action that carries risks, ranging from relatively common methotrexate side effects like nausea, vomiting, and even fatigue (which she's been dealing with already), to long-term risks of blood disorders or liver injury. Even with a plan to reconsider pregnancy when and if her RA improves significantly—perhaps in months or years—Kim will be older and potentially face greater challenges in terms of conception, successful pregnancy, and birthing a healthy baby. Whether the methotrexate works or not, she'll face another treatment decision—discontinuing methotrexate, and finding an appropriate alternative or halting treatment entirely—when she and her husband decide to pursue their planned pregnancy.
- Using a pregnancy-safe alternative medication to treat Kim's RA, so that she and her husband can pursue their planned pregnancy. Her rheumatologist told Kim that there are several alternatives, which may be used independently or in combination. They range from oral DMARDs like sulfasalazine or hydroxychloroquine, to steroids like prednisone, to injectable DMARDs like etanercept. Each of these medications carry their own variable potential to heal, and also to harm. Each also will have to be integrated with Kim's daily life in one way or another besides the obvious questions of efficacy and side effects. The oral DMARDs are relatively inexpensive, but may need to be taken twice a day. If prednisone is needed to treat a flare-up of RA symptoms, dosing might be complicated and produce short-term side effects that prevent Kim from working or taking care of her kids. Etanercept is very expensive, and because both Kim and her husband are self-employed, they have limited insurance coverage for prescription drugs and will have to pay a significant amount out-of-pocket for these medications. That could seriously interfere with their household budget and intent to save money.

- Not treating Kim's RA at all, but continuing to use pregnancy-safe medications to reduce inflammation and pain. Her rheumatologist doesn't recommend this course of action because it will leave the underlying disease unaffected, which means that it could continue to damage her joints and produce the potential of longer-term health consequences—even while the symptoms could be masked. It would allow Kim and her husband to proceed with their planned pregnancy, though. And Kim's rheumatologist has told her that being pregnant might even attenuate her RA symptoms temporarily. There is a chance that her RA could remit naturally over the course of several years, but there's no way of knowing. Kim and her husband aren't too fond of this non-treatment option, but it's on the table.

In Kim's situation, there doesn't seem to be one obviously best course of action. It is not clear that delaying their planned pregnancy and treating with methotrexate is ideal, nor proceeding with pregnancy but using a different medication. Her rheumatologist is relatively familiar with evidence about the benefits and harms of these medications, and has a good deal of clinical experience with patients choosing many different courses of action. But even that rheumatologist cannot tell Kim what to do. Kim has questions about what might happen if she chooses one course of action or another:

- If she doesn't treat with methotrexate now, will her RA permanently disable her or ever be as treatable again?
- Will "pregnancy-safe" medications potentially save her from pregnancy risks related to RA treatment, but mean that she will suffer worse RA long-term?
- How will side effects from medications make her feel during pregnancy?
- Is methotrexate that much better than the alternatives?
- Is it worth considering a really expensive medication like etanercept?
- What would it be like to inject a medication every week?
- How long will it take for any of these medications to work and for her to feel better?
- Is it worth putting off pregnancy to get her RA "under control?" Will she be able to get pregnant later? Is she risking a miscarriage, or an unhealthy baby?

Some of these questions are exceedingly difficult to answer. There may be no data, or it may be a matter of intuition. For example, even when there are data about how well medications work it is more or less impossible for "medicine" or "medical research" to answer that final question about whether it's worth delaying a planned pregnancy.

Jerry (established RA)

Jerry is a married 63 year-old retiree, with two grown children. His wife is slightly younger than him, and still working. He is a tinkerer and a woodworker—he makes knick-knacks like birdhouses for friends and family—and general household handyman. He also likes to play baseball with his young grandkids, and weekly tennis with his wife. However, he has also suffered

from RA for several years. And recently, after being relatively symptom-free (though not in remission) with methotrexate, his RA has progressed and the methotrexate no longer seems to be working effectively. Recently, Jerry has not been able to work in his workshop or around the house, or play with his grandkids. He has been stiffer than usual every morning, tired, and feeling depressed. He hasn't been active and has been eating poorly, which is doubly bad for Jerry, because he has also suffered a minor heart attack and is taking a statin for high cholesterol, a medication to lower his blood pressure, and daily aspirin—to prevent a second heart attack or stroke, which he is at higher risk for because of both his history and his RA.

Jerry's rheumatologist explained to him that sometimes methotrexate does stop working, which has upset and perplexed him since it had been working so well for so long. Suspecting a flare-up, Jerry's rheumatologist a few months ago prescribed a short-term course of prednisone. However, that didn't seem to affect the general worsening of Jerry's RA and led to the suspicion that the methotrexate that had been both well-tolerated and effective, was no longer going to be sufficient for both relief of RA symptoms and to prevent progressive joint damage. But because of the impact the medication failure has been having on Jerry's life, his rheumatologist has recommended that Jerry either add a medication to his treatment regime or try replacing the methotrexate with an alternative. His rheumatologist has presented several general options to Jerry, all of which are fairly expensive medications and new to Jerry, who is uncomfortable with starting a new medication. Among the courses of action Jerry could take are:

- Adding another injection medication that Jerry can take at home once or twice a week, such as etanercept or adalimumab. These are also expensive medications with known serious side effects.
- Discontinuing the methotrexate, which his rheumatologist thinks may no longer be helping much anyway, and trying one of the above medications alone (without methotrexate). Jerry is concerned about this option because of his historical success with methotrexate and because he had gotten used to his routine with it.
- Similar to the above options, Jerry's rheumatologist explains that there are other medications may work differently than the injectable medications adalimumab or etanercept, but which can be taken just once a month or every two months at a clinic. It would mean sitting down for a couple of hours and having an IV placed. Jerry likes that it seems as though it would be a relatively infrequent activity—better than taking two more pills every day—but does not like that it would "make him feel like a cancer patient," revealing fear of feeling overly medicalized.
- Adding a low dose of a steroid to complement the methotrexate. This would be a relatively inexpensive option, but the rheumatologist is less confident that it will work well since using prednisone to treat what they both thought was a flare-up didn't seem to have a lasting effect.
- Changing nothing, continuing to take methotrexate and carefully attending to both his symptoms and the progression of his RA.

Jerry wants to take the advice of his rheumatologist, but his rheumatologist didn't give him a single option. Because neither he nor his wife are particularly savvy with medications or medical decisions, they enlist the help of their

daughter who is a graduate school-educated biologist and more comfortable navigating these kinds of decisions. Again, there is no clear best path. Jerry has to decide how comfortable he is with the idea of taking a new medication, and if he does that, what kind of medication he is willing to try in terms of the ways in which it might interfere with his life. He wants to know what will work the best, because his ultimate goal is simply to be as free as possible from RA problems, health problems generally, and the bother of medications and side effects.

Both of these scenarios illustrate complex medical decisions, in which there is no obvious "best" course of action. In order to make an informed decision, the patients—and frankly, their providers—might reasonably be expected to ask and desire information on the consequences of particular choices before making a decision. They also underscore the very human experience at the heart of medical decision making. If their doctors were to recommend a particular course of action, they might want to know more about that course, or alternatives. What they want to know might not be easily answered during a single clinical visit.

The overloaded clinical encounter

"The perfect physician is one who knows all the science and technology relevant to medicine, who knows how to apply it skillfully for the benefit of one's patients, and, at the same time, is a wise compassionate counselor who earns his or her patient's trust and provides comfort in times of illness and need."

(Wacker, 1984)

This project's premise is that self-directed online search should yield data-driven applications—at least for those who want them. Helping those people become informed participants in their own medical decision making is one important rationale, but so is relieving pressure on the clinical encounter, when patients and practitioners meet in the proverbial clinic. The encounter is, in principle, a good place to seek answers about medications. It goes without saying that a trusted, knowledgeable physician who understands a patient's life, goals, preferences, and health status is a good partner for helping that person process medical information. However, to locate education and question-asking *only* in the clinic demands much of an expensive, time-limited , and often emotionally charged moment naturally suffused with an information and power asymmetry (the doctor knows more, and is not the person who is "sick"). It seems unlikely that all the question-asking and answering, all the information-absorption, and all the critical details of a person's health inquiry

can be addressed in a 20-minute doctor's visit—or perhaps a 45-minute visit, or even two such visits (Dugdale, Epstein, & Pantilat, 1999). The visit may be days or weeks away from the moment of first concern, and that is imbued with the stress of resulting in a clear diagnosis or treatment decision. Perhaps only in rare and ideal circumstances will a patient come prepared, perfectly articulate, with an agenda that they and their provider agree on and have time enough to cover, ensuring that all the patient's questions are asked and answered accurately—with persistent and accurate absorption of information. Many barriers exist to an ideal encounter (Schattner, 2014). More likely is a modestly successful visit in which the doctor and patient roughly agree on what to cover, and how, and the patient leaves with perhaps some questions—but not too many. Nevertheless, some queries might remain unasked or unanswered.

Similarly, locating all education in the clinic places an enormous burden on the clinician's shoulders. While one's doctor is a highly trusted source of medical information (Hesse et al., 2005), it's unrealistic to expect any clinician to be exhaustively knowledgeable about every topic that a patient might graze—especially considering the "doorknob phenomenon." (Imagine the patient just about to leave, putting on their coat, saying "oh, by the way, there's one more thing I wanted to ask about…") It is unreasonable to expect every provider to be intimately familiar and up to date with the vast body of medical evidence—and to be able to contextualize it, summarize it, and clearly communicate it to the patient who's just asked an important question on their way out the door. Even if the patient arrives with an articulate, well-researched question, and is prepared to have a discussion about it, there may not be enough time, and the *provider* might not be prepared to discuss it. Furthermore, clinicians are expected not only to provide excellent, high-quality, and medically sound care, but to provide emotional support too—an extant, challenging burden.

Time before, between, and after visits is crucial. These moments are spaces for absorption, consideration, and independent research. There may not be the presence of a knowledgeable partner (a trusted healthcare provider), but there might be less pressure, and a friend or family member instead. Querying a search engine about one's own health may be an informative, but also very private and emotional experience. We are free to express or ask about things that we might not feel comfortable divulging in other circumstances. Out of the clinic, questions can be asked at a pace under the patient's control, and new avenues can be explored at will. People can make use of multiple kinds and sources of evidence—from the high quality evidence that underpins modern medical practice to the unverified and anecdotal—and learn by way of whatever modalities suit them best—graphical, interactive, textual and narrative, aural, and by way of videos and animations. Clarification can be sought without fear of interrupting a physician during a time-limited clinical encounter. A "better" Dr. Google, prospectively, is one that supplies resources that are based on evidence that patient and practitioner might both value highly.

Online health search and trends in medical decision making

Though RA may only affect just over 0.6% of U.S. adults (Helmick et al., 2008), there may be more search interest in it than for more prevalent chronic conditions (see Chapter 2). Online health search today is dominated by results from Web sites like WebMD—primarily static, non-interactive, narrative, and text based. That is particularly true for searches about RA medications, which ordinarily have no details about improvement outcomes and limited data about incidence of side effects. Some of the richest user experiences and data-driven resources come from RA drug manufacturers, which as paid advertising results may appear "disguised" to an ordinary person. Rarely, resources *are* data-driven and may even include patient-reported data—such as with the Web site PatientsLikeMe. Data from clinical trials and systematic reviews of the best available research on RA medications is much harder to come by in "consumer-facing" health resources. (Chapter 2 has more detail on online search for RA medication information, and commonly used Web sites.)

Meanwhile, *non*-health search and Web applications are often data-driven. Products like Google Maps for navigation and Kayak or Hipmunk for travel planning feature visualization-heavy UIs that facilitate interaction with rich multivariate datasets, from which people make meaning. With these services, one can easily find a public transit route to the airport, and the cheapest flight with the shortest connection. Admittedly, the data powering such resources are much less complicated to interpret than outcome data about RA medications. Nevertheless, they are indicators that well-designed UI can help ordinary people interact with large volumes of data.

While online health resources appear to "lag" behind non-medical counterparts, the practice of medicine itself is changing rapidly. In addition to some of the aforementioned changes around sharing clinical trial data which might directly affect availability of evidence useful for the *creation* of data-driven online health resources, the increasing engagement of patients themselves in decision making may provide an impetus for more such resources. Alongside the movement for practitioners to adopt evidence-based medicine (EBM)—in which the best available evidence from clinical trials or clinically applicable epidemiological studies provide the rationale for medical decisions (Eddy, 2005; Guyatt et al., 1992)—so-called patient-centered practices that shift focus from treatment of *disease* to care of a *person*, are equally being promoted and taken up. For example, shared decision making (SDM), in which patients are invited to participate in medical decisions in concert with their clinicians (Charles et al., 1997; Stiggelbout et al., 2012), which is a suite of concepts around communication, education, and decision making that is intended to respect the individual needs, preferences, and values of the patient. Increasingly, there is a recognition of the need for practitioners and their patients to find common ground on care plans—especially those that are truly manageable to patients. Patient-reported outcomes—those which patients measure themselves directly, on outcomes that are perceptible to them—are becoming more important in RA and other conditions (J. R. Kirwan, Newman, Tugwell, & Wells, 2009). Clinical practice guidelines for RA, which outline standards of care and make recommendations around diagnostics and

treatment, promote both evidence-based practice *and* shared decision making (Singh et al., 2016; Smolen et al., 2013).

There is a good deal of research on patient-oriented decision aids (DAs), which are artifacts or tools designed to help people understand and make informed medical decisions in line with their needs and values (Stacey et al., 2012), and which may be used in consultation with clinicians. DAs based on the very same evidence practitioners use might help with patient-centered decision making, as might online resources–like those envisioned by this thesis project–developed with the same evidence (and more), but designed explicitly for use at home, to complement or ultimately replace outmoded, static resources.

The RA Choice decision aid

The RA Choice decision aid (Barton et al., 2014) forms the basis of the prototype that is the primary output of this thesis project. Indeed, the first version of a user-facing application built with the prototype was a digital version of RA Choice (thesis.merges.net/ptda). This decision aid (DA) was designed for use when a patient with established RA for whom at least one trial of methotrexate had been ineffective. It is similar to Mayo Clinic decision aids (see Chapter 2) in that it presents medications in a similar visual position across 5 "issue" cards (see Figure 1-1). The title of the article describing the DA–*The design of a low*

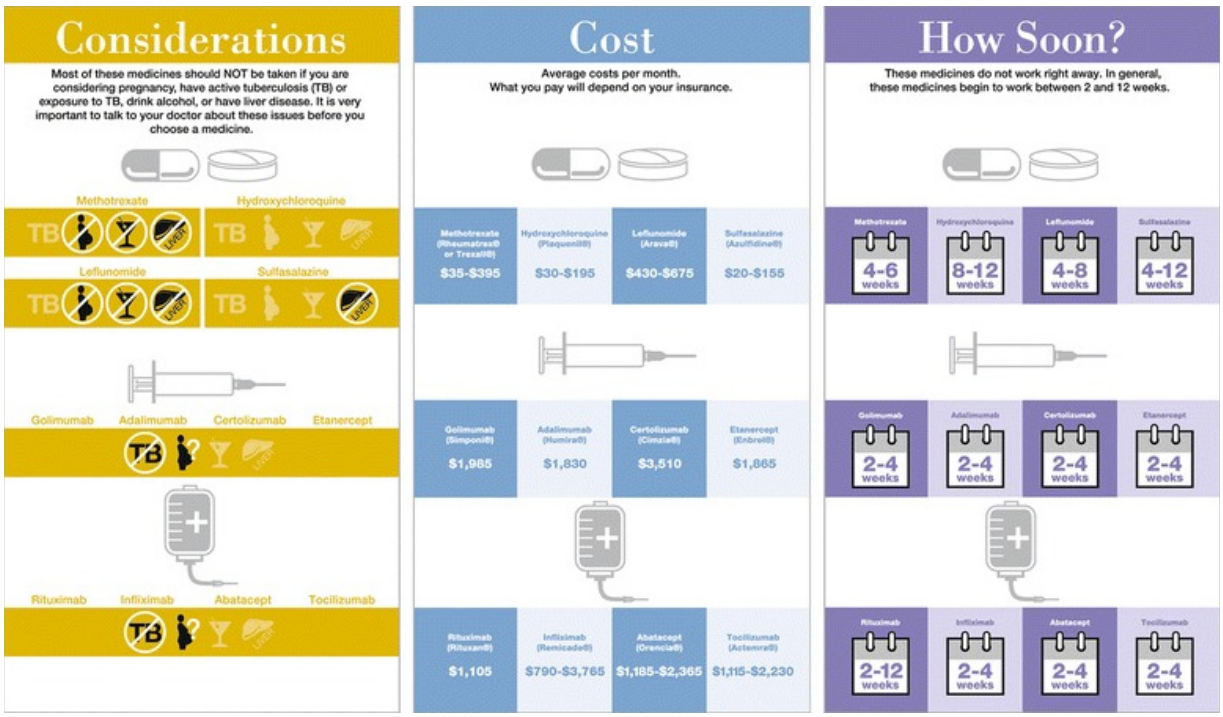


Figure 1-1
RA Choice decision aid. These are 3 of the 5 issue cards. The remaining cards are about side effects and frequency of administration.

literacy decision aid about rheumatoid arthritis medications developed in three languages for use during the clinical encounter—explicitly addresses the rationale underlying the design. Barton et al. point out that "disparities in RA outcomes and reports of poor shared decision-making communication among vulnerable populations" bespeak a need for tools to hopefully address such disparities. Like the Mayo Clinic DAs, there is an emphasis on graphics and plain language.

The DA was developed using a rigorous user-centered process with constant input from designers, clinicians, and patients of multiple literacy levels in the service of a DA that is not necessarily self-contained, but instead expressly for the purpose of facilitating a conversation in the clinic (Barton et al., 2014). The resulting set of 5 issue cards and their content, including the dimensions of the "considerations" card in Figure 1-1, is the product of extensive literature review and synthesis, and design iteration with patients.

Since this DA is the basis of the prototype—it supplies the set of 12 medications, and some of the key dimensions along which they can be "filtered"—it is unfortunate that some of its low-literacy design qualities have been abandoned in the present state of the prototype. However, an "enhanced" version of the RA Choice DA, preserving the initial design intent and low-literacy features but elaborating on it with the capacity to explore evidence for each medication, could relatively easily be built using the prototype as a platform. Nevertheless, the question that prompted the development of the prototype is: What evidence might complement this DA? And what stands in the way of incorporating that evidence about these 12 medications? Barton et al. write that they experimented with the design of a card with visualizations of efficacy but ultimately rejected it due to lack of head-to-head trials for all medications, and that such a card could mislead patients. Those challenges are front and center in this thesis project.

Factors that shape or limit access to medical evidence

"The paper is not the discovery. The data are the discovery."

Dr. Ruben Abagyan, Professor of Pharmacology, University of California San Diego at a presentation to the Council of Science Editors annual meeting in 2010.

There are few sources of specific medical data (e.g. an estimate of how effective aspirin is for pain relief after a particular injury) accessible to the general public. Desirable real-world data (not from clinical trials sponsored by drug companies) about real-world side effects—such as their magnitude, frequency,

onset, and duration—is difficult or impossible to find. Equally difficult to find are data about patient satisfaction with intervention options. Numerous barriers separate an interested patient and medical evidence of almost any sort. Most people do not have the means to read medical literature, which is mostly published in costly, access-controlled scholarly journals (though there are a few prominent open access peer-reviewed journals). The literature is written in domain-specific jargon suitable to its audience—for the most part, other researchers. Research findings may be only indirectly applicable to real-world decision making, one reason being that it can be difficult to measure and communicate information of value to patients for medical decisions. In many cases, even published research that is directly applicable, comprehensible to someone with adequate health literacy, and freely accessible is only partially complete. Instead of publishing data or conclusions in a systematic way, researchers may publish only a few of the most salient or novel findings, leaving the full suite of data in a desk drawer. Some of these issues are germane to this project and explored in the prototype and discussion on subsequent pages. Although this is the context from which the prototype springs, a full discussion of the social and technical complexities preventing wider access to medical evidence is far beyond the scope of this project.

In many cases, a *summary* of the best available evidence *is* readily found online or in patient educational literature. For instance, none of the most commonly prescribed antidepressants (SSRIs, SNRIs, bupropion and a few others) has been found superior for the treatment of depression. This very general summary is repeated in many consumer resources. WebMD says "no antidepressant works better than another" (WebMD, n.d.) and the Mayo Clinic Depression Medication Choice decision aid says "the antidepressants presented in this decision aid all work the same for treating depression" (LeBlanc A, Herrin J, Williams MD, & et al, 2015). However, that is not the whole picture, parly because it focuses on just one broad outcome. If one wishes to learn in some detail *how* effective these medicines are, *for whom* they may be most effective, or estimates of potential harmful effects, there are fewer accessible resources. Slightly richer plain language summaries of evidence, such as those found in the first few pages of systematic reviews published by the Cochrane Collaboration, are published by companies who sell access to them by subscription or expensive one-time payments (often $30 per publication). Clinical trials, real-world patient experiences, epidemiological studies of data from patient registries, and other types of research are all valuable sources that enrich medical knowledge. Rarely are machine-readable systematic data published, though; in the majority of literature, findings are communicated in written language and in tables. For researchers, educators, designers, and others to be free to turn the *findings* that might shed light on details into *resources* that do—whether written, visualized, interactive, paper-based, digital, jargon-filled, jargon-free, low-literacy, animated, static, dynamic, or otherwise—the *data* underlying these publications must be shared and machine readable. With the data, one can choose to write a paragraph summary, perform a statistical analysis, *or* build an app.

Data sharing

As a matter of discussion and policy, the imperative of sharing clinical trial data has ascended in recent years. In early 2016, the editors-in-chief of many of the world's most prominent medical journals published an editorial outlining

requirements for data sharing and proposing that authors be *required* to "share with others the deidentified patient data (IPD) underlying the results presented in" submitted articles, "including tables, figures, and appendices or supplementary material" (Taichman et al., 2016). This follows similar calls from institutions like the National Institutes of Health (Hudson KL & Collins FS, 2015), the WHO (World Health Organization, 2015), and the British Medical Journal (BMJ) (Loder & Groves, 2015). These proposals—and in some cases, new rules—are an enormous accomplishment, especially given the inertia of medical research as an institution. They typically describe such sharing as an ethical imperative, with a scientific rationale: Original data are necessary so that other researchers can perform analyses on the widest available set of data about interventions—not just "positive" results—in order that dissemination bias can be attenuated, and the truest state of the science can emerge (Moorthy, Karam, Vannice, & Kieny, 2015). Clinical trial data are among the most important data for understanding the true effects of medical interventions, but they are not unique. Data underlying systematic reviews, post-hoc epidemiological studies, post-marketing adverse effects surveillance—truly, *any* kind of research—are potentially useful for painting the most complete picture possible. Data sharing, however, is beset by so many social and technical issues it remains an aspiration. A 2015 publication on the subject by the Institute of Medicine (IOM) of the National Academies of Science (Institute of Medicine, 2015) begins with this unequivocal statement: "Patients and their physicians depend on clinical trials for reliable evidence on what therapies are effective and safe." However, even the simplest requirements for sharing—such as submission of a lay-language summary of results, required by U.S. law since 2007—are often unmet challenges. Quoting Saito and Gill (2014), the IOM authors write that "a 2012 study found that the results of 30 percent of 400 clinical trials had neither been published nor reported on ClinicalTrials.gov 4 years after study completion." Just a few of the publication's 300-odd pages are devoted to the exceedingly complex technical challenges of data sharing; it is almost entirely about the social, political, and regulatory practices that may be necessary to implement such a vision. While we wait for the social machinery to enact its de-biasing science, it is worth investigating other strategies to extract existing machine-readable findings that can be developed into useful patient-centered resources.

The data: Machine-readability

If data are not machine readable—stored and encoded in a format useful to computer programs—they effectively cannot be used to develop interactive, tailored, digital resources for the general public, not to mention decision support tools that might equally benefit clinicians. Consider the data sharing declarations of 2016 in light of this description from one of the first scholarly articles about evidence-based medicine from 1992, in which a junior medical resident is treating a 43-year old man who experienced a grand mal seizure, under the heading "The Way of the Future:"

> The resident asks herself whether she knows the prognosis of a first seizure and realizes she does not. She proceeds to the library and, using the Grateful Med program, conducts a computerized literature search. She enters the Medical Subject Headings terms epilepsy, *prognosis*, and *recurrence*, and the program retrieves 25 relevant articles. Surveying the titles, one appears directly relevant. She reviews the paper, finds that it meets criteria she has previously learned for a valid *investigation*

> of prognosis, and determines that the results are applicable to her patient. The search costs the resident $2.68, and the entire process (including the trip to the library and the time to make a photocopy of the article) took half an hour.
>
> The results of the relevant study show that the patient risk of recurrence at 1 year is between 43% and 51%, and at 3 years the risk is between 51% and 60%. After a seizure-free period of 18 months his risk of recurrence would likely be less than 20%. She conveys this information to the patient, along with a recommendation that he take his medication, see his family doctor regularly, and have a review of his need for medication if he remains seizure-free for 18 months. The patient leaves with a clear idea of his likely prognosis.
>
> (Guyatt et al., 1992)

Guyatt et al. say that this kind of practice, "which involves using the medical literature more effectively in guiding medical practice, is profound enough that it can appropriately be called a paradigm shift" per Thomas Kuhn's notion of a scientific revolution. Sadly, however, their description of the future—reading high-quality, clinically relevant *articles*—barely describes the state of the art a quarter-century later. Their humble technical vision, written just as the first laptop computers appeared, before widespread Web access and Internet-enabled mobile devices, has not entirely come to fruition, while in other domains it has been surpassed by mind-boggling advances in data-driven applications. Summaries of medical evidence do exist, and they are being used to inform clinical practice. Partly because of the social and technical issues that have heretofore limited data sharing—to say nothing of the generation of *relevant* findings—there are few sources of clinically and patient-relevant machine-readable medical evidence that can be, or are, used for data-driven resources. Enormous effort has been expended to build computer systems—through artificial intelligence, machine learning, and natural language processing research of all sorts—that might sift through vast quantities of data, such as unstructured findings hidden in the text of the scholarly medical literature, and extract useful meaning. Despite the popular press praising IBM's Watson project, no breakthrough has yet emerged that can substitute for well-encoded machine-readable data. Researchers have produced intriguing findings when Watson machine learning technology was applied to tens of thousands of abstracts of basic research articles about a kind of human protein called kinases, in order to predict which kinases modify another protein, p53 (Spangler et al., 2014). It is unclear whether these methods would produce valuable results on a much smaller dataset, to help guide a decision about which RA medication is appropriate. State of the art evidence-based medicine resources from the Cochrane Collaboration that *can* inform clinical decisions as Guyatt et al. envisioned in 1992 *do* exist on paper, and in digital paper (PDF) format—but even that group has not produced a machine-readable data solution, only presentations and blog posts that gesture at the outlines of such a solution (See "Cochrane Linked Data," Cochrane Collaboration, n.d.-a).

In order to get around the absence of established standards and practices for machine-readable findings from medical evidence, the prototype demonstrates ways in which findings can be encoded by a modestly knowledgeable reader of clinical trials, systematic reviews, and other sources. Discussed later, the approach is limited, but it is straightforward and reproducible, and in that

it can be used to develop a wide variety of data-driven resources for both clinicians and patients, it is worthy of investigation and as a demonstration.

Translation and design

If the findings from medical research *were* machine-readable and shared, would they be applicable and understandable to people making medical decisions? It depends. In many cases, research is not relevant to an individual's life—about outcomes that may not be of decision-making value or are one step removed. In other cases, the information *is* likely valuable but may still require translation to make it understandable. Where research has been standardized—for example, on a set of relatively consistent clinical trial outcomes such as the OMERACT outcomes for RA medication research (Boers, Kirwan, & Tugwell, 2014)—it is certainly easier to harmonize multiple sources or present them side-by-side, as a first step. Even still, a great deal of translation is necessary from *data* to visualization and text, from jargon to plain language, from dataset to Web application, and so forth. With machine-readable evidence, the disciplines concerned with this translation, from health education to programming to data visualization to UX design, can focus their efforts on helping people make meaning from it, rather than manipulating or extracting findings from it. In this way, the full promise of evidence-based medicine is predicated on machine readable data with which specialists can work to translate and design for practical use.

Non-scholarly sources

Although scholarly research—in the form of data from clinical trials, meta-analyses, epidemiological studies, qualitative studies, and so forth—are critically important, they are not the only kinds of evidence that may matter to people seeking medication information. Real-world medication experience data, patient testimonials, doctors' opinions, prescription drug formularies, real-world cost surveys, and other kinds of evidence may be desired by patients. Similar social and technical barriers stand between these kinds of data and their use in online health resources. They must be properly shared or available, encoded, and translated (or put in context).

Since this thesis project could not hope to tackle the social, political, and economic complexities that shape access to medical evidence, it instead centers on encoding data and demonstrating translation and design on top of those data. In the spirit of sharing machine-readable data, the source code and all data presented in this thesis and in the prototype are online and available for others to use, echoing one of the project's primary objectives.

Project objectives

Against the backdrop presented thus far emerged this thesis project, which explores practical problems in how best to make medical evidence available and understandable to the general public. It demonstrates in a limited way how certain contemporary technologies and UX design can be used to encode, communicate, and facilitate interactive exploration of research about medications that treat RA. The project is about those medications specifically, but conceptually the work transfers to other medical domains, including treatment for other chronic conditions.

The practical problems center on these three themes:

- A lightweight technological architecture that can accommodate relevant medical evidence, including arbitrary new data.
- A basic automated translation and harmonization system that helps the general public make meaning out of medical evidence, such as by re-describing idiosyncratic outcome measures in plain language.
- Appropriate design (user interfaces and visualizations) to make the data understandable.

As mentioned in the introduction, merely discussing the shape of systems that could hypothetically address access to medical evidence is insufficient. Just as the (probably apocryphal) Martin Mull quote goes, "writing about music is like dancing about architecture," so writing about design is like dancing about architecture. Only a built artifact—even if it fails in many ways—can stress the ideas in practice. The prototype—the design project—is the discovery, if you will—a concept not dissimilar to the divide between journal articles and the underlying data.

This project's overriding objective was to create an open-source interactive, expandable, malleable, data-driven tool for the general public to explore medical evidence about rheumatoid arthritis medication options. Its further objectives were to:

1. Demonstrate an end-to-end technical architecture, based primarily on open-source technology, for the creation and deployment of data-driven medication information applications or Web sites.
2. Demonstrate some strategies for automating the difficult work of translating from opaque evidence (e.g. jargon-filled, or using oblique or surrogate outcomes) to understandable forms (of language and visual design).
3. Demonstrate and discuss tradeoffs in user interface design choices and rationale, as they relate to making medical evidence accessible.

It should be noted that this project was developed by one person (with feedback from academic supervisors), absent participation from patients, practitioners, or programmers. Due to time and research constraints, *evaluation* of these objectives' success in practice must be relegated to future work. However, each of these audiences can now be engaged to build on the prototype.

Audiences

The prototype has three audiences. It is intended to demonstrate different notions of value to each respective audience, and may be seen as the basis for future work in each of these thematic areas.

Patients and caregivers

For people with RA or their caregivers, the prototype demonstrates how a publicly available Web app could provide data to help them learn more about medication options. While it cannot answer difficult questions—such as those presented in the earlier scenarios—it does illuminate details on various treatment options. The prototype might show this group that there are ways to learn from data about medications, other than reading academic articles or static pamphlets or Web pages.

Doctors, nurses, pharmacists, researchers, or other contributors from the medical community

These potential users might recommend this kind of application for patient education or use in SDM. More importantly, they could contribute to it by encoding findings from research simply by reading articles and adding rows to a spreadsheet. It is intended to show this group that medical evidence can be designed effectively for the general public, and that free, lightweight, and easy to use technologies like spreadsheets are sufficient for encoding and communicating medical evidence—as opposed to expensive, bespoke, and difficult to maintain custom-programmed computer systems. When new data are added to this prototype, they show up immediately. Further, it is a platform on which infinite applications or visualizations can be built. One person built this prototype; it can be much more with the contribution of specialists in the medical community.

Designers, programmers, and educators

For these users, the whole "stack" of technology—from data encoded in spreadsheets, to the translation layer and controlled vocabulary in the middle, to the ultimate user interface—may be of interest. It can be used as a model for encoding other kinds of medical evidence, or for other conditions, or the UI components in the prototype can be re-used in other, related projects. Because it is open source, this audience can see how the whole system works, can replicate it, and pick and choose parts that are valuable for their own work.

Chapter 2

Context, rationale, and related work

Online health information search

Before and after clinical encounters, the information found Googling becomes a part of the milieu of knowledge involved in medical decisions—especially relevant when patients are making complex medical decisions. Online search is convenient, can reinforce concepts or knowledge imparted during clinical encounters, and supply details that were missing. In 2012, 59% of the U.S. adult population searched online for health information and 43% of adult Internet users specifically sought information about a certain medical treatment or procedure (Fox and Duggan, "Health Online 2013", 2013). Just *what* they search for varies by health condition. Since nearly 50% of adult Americans have been diagnosed with at least one chronic non-mental health condition (Ward et al., 2014), it is illuminating to look at how RA-related searches (and possibly consequent informational needs) compare to other conditions. Looking at Google query trends is not unknown to medicine (Harsha, Schmitt, & Stavropoulos, 2014). Search volume data are trade secret, although Google provides coarse estimates of monthly average query volume through its freely available AdWords Keyword Planner application (Google, n.d.).

One way of examining relative interest in online information about certain medical conditions is to obtain estimates of search activity for general search queries (or *keywords*) related to those conditions. Those estimates can be contextualized in prevalence data. Conditions with a higher number of estimated searches per person (i.e. patient) per year suggest more interest, and therefore a greater need amongst that population—although it must be noted that online searches for terms like *depression* and *rheumatoid arthritis* are not necessarily by or for people diagnosed with those conditions. Nevertheless, assuming a similar distribution of topical patient and non-patient queries for these very general terms, it is instructive to look at these searches-per-patient as a rough indicator of interest, understanding that their reliability is limited.

Table 2
Estimated chronic condition search volume. These are estimates for broad search terms only, and therefore a very rough guide to search interest.

Condition	Search term(s)	**Yearly estimated U.S. Google search volume** Based on February 2016 average	U.S. adult prevalence	Estimated U.S. Google searches per patient per year
asthma		**4,417,200**	6.69%[1]	**0.27**
	asthma	1,620,000		
major depression		**4,334,400**	3.40%[2]	**0.52**
	depression	3,612,000		
	major depression	397,200		
	major depressive disorder	325,200		
COPD		**4,417,200**	6.07%[3]	**0.29**
	copd	2,952,000		
	chronic obstructive pulmonary disease	145,200		
	emphysema	1,320,000		
diabetes		**7,644,000**	8.86%[4]	**0.35**
	diabetes	3,612,000		
	type 1 diabetes	1,620,000		
	type 2 diabetes	2,412,000		
heart failure		**2,899,200**	2.06%[5]	**0.57**
	chf	594,000		
	heart failure	325,200		
	congestive heart failure	1,980,000		
high blood pressure		**6,012,000**	29.10%[6]	**0.08**
	hypertension	1,620,000		
	high blood pressure	1,980,000		
	blood pressure	2,412,000		
osteoarthritis		**3,732,000**	12.33%[7]	**0.14**
	arthritis	2,412,000		
	osteoarthritis	1,320,000		
rheumatoid arthritis		**4,572,000**	0.61%[8]	**3.05**
	ra	1,620,000		
	rheumatoid arthritis	2,952,000		
HIV/AIDS		**3,300,000**	0.49%[9]	**2.75**
	hiv	1,980,000		
	aids	1,320,000		

1. Current Asthma Population Estimates (Centers for Disease Control and Prevention, 2014).
2. Current Depression Among Adults–United States, 2006 and 2008 (Centers for Disease Control and Prevention, 2010).
3. Chronic obstructive pulmonary disease among adults–United States, 2011 (Centers for Disease Control and Prevention, 2012).
4. Number of Adults - Diagnosed Diabetes - Diabetes DDT (Centers for Disease Control and Prevention, n.d.)
5. Heart Disease and Stroke Statistics (Go et al., 2013)
6. Hypertension among adults in the United States (Nwankwo, Yoon, Burt, & Gu, 2013)
7. Based on Estimates of the prevalence of arthritis and other rheumatic conditions in the United States–Part II (Lawrence et al., 2008)
8. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States–Part I (Helmick et al., 2008)
9. Prevalence of Diagnosed and Undiagnosed HIV Infection–United States, 2008-2012 (Centers for Disease Control and Prevention, 2015)

Table 2 shows general search query terms, average yearly search volume, estimated U.S. adult prevalence data from sources cited by the Centers for Disease Control and Prevention, and relative interest (searches per patient) for several chronic conditions. There are is more than an order of magnitude difference between relative interest in very widespread conditions (like hypertension and osteoarthritis) and RA. From these data there is no way to infer *why* there is a difference between conditions, but it may be that some may produce unfamiliar symptoms, are treated with unfamiliar medications, or have an unfamiliar course. All may motivate someone to search online in an effort to understand these symptoms, treatments, and outcomes. Common and burdensome conditions (in terms of symptoms and treatments) like diabetes, chronic obstructive pulmonary disease, and congestive heart failure have higher relative search interest. HIV/AIDS, similar to RA in terms of prevalence, has a similar relative search interest–again noting that people may search for these general search terms even if they do not suffer from HIV/AIDS, which may overestimate search interest among people living with the condition. Repeated searching could be another explanation. In all these cases, searches for specific treatments were omitted; search interest for a selection of medications that treat RA can be seen in Table 3.

Table 3
Estimated RA-related medication search volume.

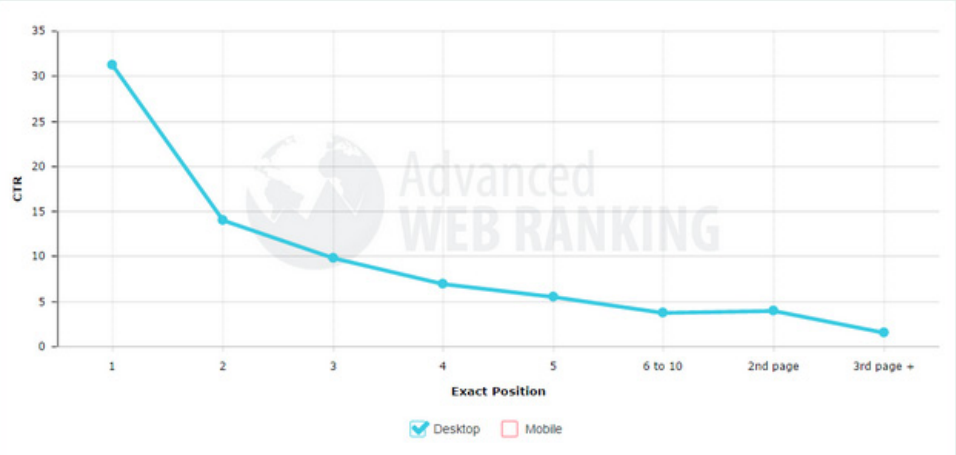
Medication/group	**Estimated yearly U.S. Google search volume** Based on February 2016 average	**Estimated yearly U.S. Google searches**
Methotrexate	1,320,000	**1,332,000**
Rheumatrex	12,000	
Humira	888,000	**985,200**
Adalimumab	97,200	
Hydroxychloroquine	397,200	**991,200**
Plaquenil	594000	
Rituximab	217,200	**483,600**
Rituxan	266,400	
Infliximab	145,200	**631200**
Remicade	486,000	
Etanercept	79,200	**476,400**
Enbrel	397,200	
Sulfasalazine	325,200	**378,000**
Azulfidine	52,800	
Leflunomide	177,600	**274,800**
Arava	97,200	
Tofacitinib	52,800	**270,000**
Xeljanz	217,200	
Abatacept	43,200	**188,400**
Orencia	145,200	
Tocilizumab	43,200	**140,400**
Actemra	97,200	
Certolizumab	15,600	**134,400**
Cimzia	118,800	
Golimumab	22,800	**102,000**
Simponi	79,200	
Anakinra	34,800	**63,600**
Kineret	28,800	
biologics	97,200	**192,000**
biologic	79,200	
biologic drugs	15,600	
DMARDs	34,800	**87,600**
DMARD	52,800	
Total for all drugs		**6,730,800**

Looking more closely at RA medications, the Google AdWords Keyword Planner suggests that there are approximately 6.5 million queries per year in the United States for the some of the most commonly used DMARDs, including methotrexate, the most well-known biologics, and tofacitinib. These estimates are based on a limited set of queries for each medication–the generic name and one brand name. An additional estimated 280,000 queries per year are categorical in nature, representing searches for the general type of medication in question–DMARDs and biologics. Estimated volume for more specific terms like *disease-modifying antirheumatic drug* and *TNF inhibitor* was very low (around or under several thousand queries per year) and excluded from this table. It must be noted that many of these medications are used to treat multiple conditions, including cancer, malaria, psoriasis, Crohn's disease,

among others. Not all search interest in these medications is related to RA. Nevertheless, since the medications are frequently used in RA treatment, these estimates provide a reasonable signal of online interest–whether on the part of patients, caregivers, or even practitioners.

While search volume estimates provide a signal of interest, they do not speak to the kinds of resources people find. First, it is instructive to examine the distribution of clicks on individual search results–called click-through rate or CTR. Unsurprisingly, people do not click equally often on all search results. A Google search for methotrexate on February 8, 2015 produced "about 7,850,000 results" ("methotrexate - Google Search" n.d.), which at 10 results per page is hundreds of thousands of search results pages. Authoritative CTR data is not available from search engine companies like Google, as it is trade secret information, and click-through rates differ depending on the particular query and results–for example, one would expect an exceptionally high CTR on the Facebook home page result for the query facebook. However, the company does provide limited data to Web site owners on queries, click-through rate, and average position in search results. One informal, non-scholarly analysis of thousands of such Web sites' data by Petrescu produced the broad estimates in Figure 2-1 of click-through rates for organic (unpaid, non-ad) result positions (Petrescu, 2014). It suggests that about 50% of users click on one of the first three results, and 70% on one of the first ten. Some 23% of queries result in no organic, unpaid clicks. In those cases, users are assumed to have abandoned the search, refined it, clicked on a paid ad, or perhaps found the answer in a search result "snippet" or summary directly on the search results page. In any case, although the estimates are rough, they indicate just how important the first few search results are.

There is no guarantee that search results will be the same from search to search, or from person to person. Which particular results appear on a given search result page in response to a query varies with location, personal search history, algorithm changes, and other factors. However, searches in a non-personalized (i.e. fresh or so far anonymous) browser window can produce



Position	1	2	3	4	5	6 to 10	2nd page	3rd page+
CTR	31.24	14.04	9.85	6.97	5.50	3.73	3.99	1.60

Figure 2-1
Estimated click-through rate (CTR) on Google search results. Screenshot from from Petrescu, P. (2014).

a reasonable approximation of typical search results. The first few results for queries like *dmards, methotrexate,* and *enbrel* are representative of the kinds of online health resources most searchers will interact with. WebMD, Wikipedia, Drugs.com, and pharmaceutical manufacturer resources rank highly.

Online "consumer" health resources

A few resources dominate the first page of search results when people query for medications that treat RA. These are the resources that most people will encounter, because of the tendency for searchers to click through to top results first. Most—like WebMD and its affiliate Web sites, Drugs.com, and Mayo Clinic—feature static prose, paragraph upon paragraph of text. In many cases these resources include restatement of pharmaceutical manufacturer-provided monographs, or text licensed from a medical content provider like Micromedex or First Databank. Because they rarely include data-driven UIs for exploring and querying medical evidence, they lag behind online resources in other domains. They are a fixed perspective on their underlying evidence. Some resources *are* different, either in the quality and type of information they provide, or in that they *are* data-driven resources—but these are rare. Following is a brief survey of the most commonly found resources, and a few that are less common but related conceptually to the prototype. Each is summarized and cursorily critiqued to emphasize the outsized role they play, and to illustrate how deliberate design decisions shape medication information presentation and the ways in which people can and cannot interact with such information.

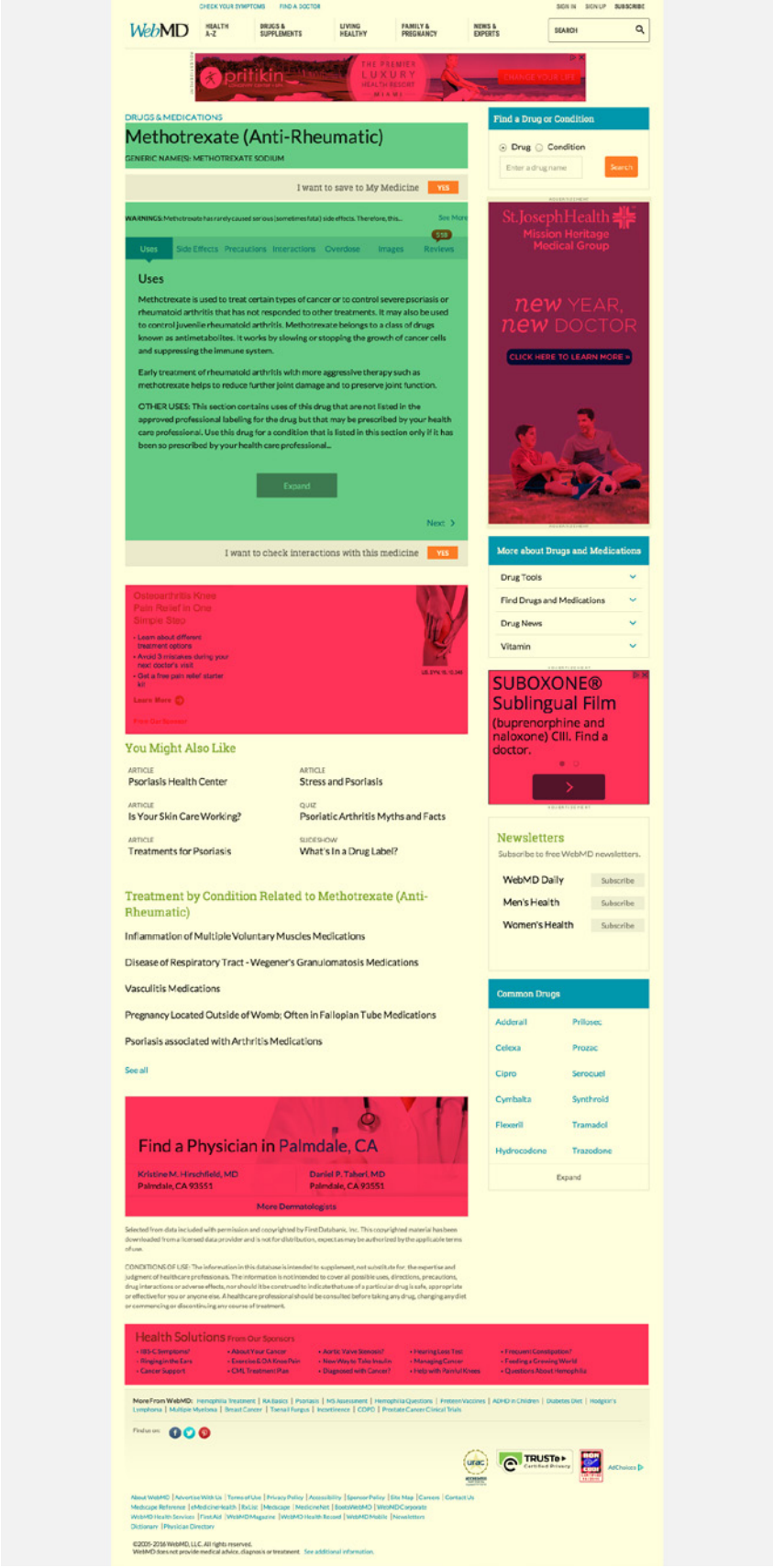
WebMD

webmd.com

WebMD is a well-known online health information resource. It features uncountable articles, slideshows, and tools about medical conditions, tests, interventions, general health and well-being, among other topics. A Google search on February 11, 2016 for the term *site:webmd.com* (which seeks all pages indexed by Google on WebMD's Web site) returned "about 16,200,000 results." Compete, an online tool that estimates visitors numbers to Web sites, estimated that more than 26 million unique American desktop computer users visited WebMD in December 2015 ("webmd.com"on Compete, n.d.). It is undoubtedly an expansive and widely used resource.

Figure 2-2 shows WebMD's page about oral methotrexate. The screenshot was taken in a browser with no browsing history or cookies. Of the usable visual space (highlighted in yellow), about 25% is devoted to advertising—for

Figure 2-2
WebMD's general methotrexate information page. Methotrexate-related content is highlighted in green. Advertising content is highlighted in red.



products as diverse as a medical practice, a luxury health resort, an injection medicine for knee pain from osteoarthritis, physician listings, a sublingual synthetic opioid pain reliever, and sponsored links to WebMD pages authored by companies like Monsanto, Pfizer, and Bristol-Meyers Squibb. Just over 15% is devoted to methotrexate-related content; so little that repeated clicks are necessary to fully consume the available information, each resulting in a new page load, and new advertising. The remainder of the space is empty or features elements that navigate the visitor to other areas of WebMD. There is valuable information about methotrexate on this page (in prose), but the design choices may belie the page's true purpose. On an RA-specific methotrexate page ("How Does Methotrexate Treat Rheumatoid Arthritis?" WebMD, 2014) more obviously pertinent information is presented on a similarly divided page. Methotrexate is described as "one of the most effective medications to treat rheumatoid arthritis," but no data are supplied to contextualize it. Similarly, benefits are described in a deterministic fashion—for instance, "it will help ease symptoms like joint pain, fatigue, redness, and swelling." A few common side effects are listed, without any details about magnitude, frequency, or duration. Sources used to write the article are listed, but this is a typical, static, non-data driven resource. While further inquiry into advertising and other WebMD content is invited by hyperlink, there are no options to tailor the information or compare methotrexate to other treatments.

General resources like WebMD, with lengthy prose, stand in stark contrast to many decision aids—artifacts explicitly designed to facilitate medical choices, which will be discussed subsequently. Decision aids often feature brief and straightforward language, pictures that communicate information or quantitative data, and visual organization that emphasizes discrete intervention options or information dimensions. WebMD and similar resources occasionally use bulleted lists, tables, or illustrations, but most critical data is peppered among long sentences, in multi-paragraph and multi-page articles.

Although WebMD has been singled out for this analysis—because it is frequently the top search result—it is representative of the status quo in online health resources. Hereafter, descriptions of similar resources will be shorter and less in depth.

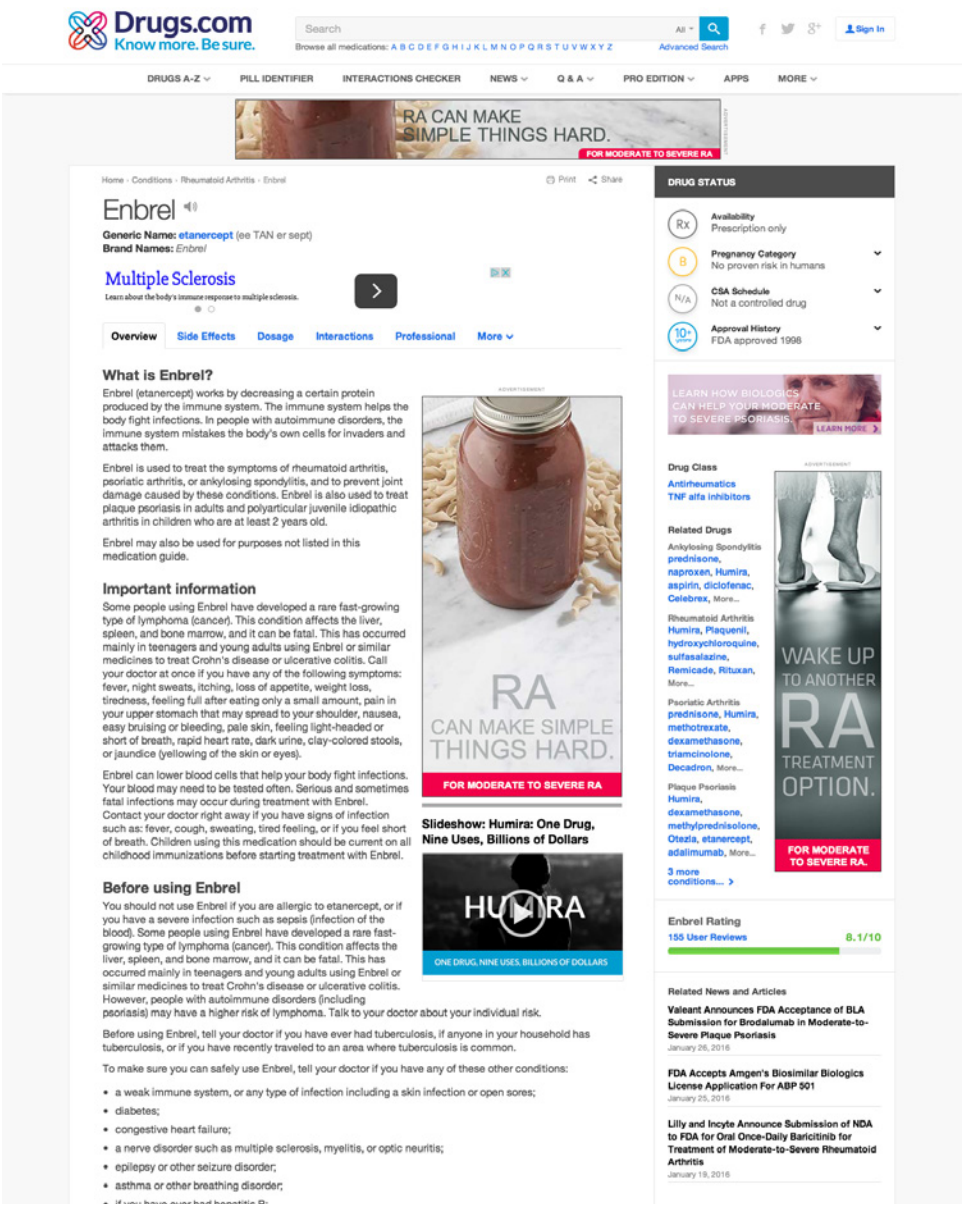
Drugs.com

drugs.com

This popular medication information Web site is not dissimilar from WebMD. Figure 2-3 shows its page for etanercept, a biologic DMARD. It is entirely textual, and features mostly general information that would ordinarily be covered in a clinical visit—contraindications, for instance. Side effects are presented with no sense of frequency or magnitude, although serious reactions are called out. There are no data about efficacy, and besides administration information very little about impact on daily routine. The content is licensed from a third-party provider of general medication information, Cerner Multum.

Since contraindications and risks are promoted on this page, it is worth noting that if someone has a specific concern (planning to become pregnant, or infected by hepatitis B, for example) the only way to learn whether etanercept is safe is to read through the content. There is no button that one can click to

Figure 2-3
Drugs.com's general etanercept information page.



indicate a concern, and get at least a preliminary but straightforward answer. Instead, a good deal of literacy is required and an investment of time to process each sentence of dense medical information—including much that is probably spurious or unnecessary. This may be considered the "kitchen sink" of text approach to medication information: Assemble it all into paragraphs, and let the patient sort it out.

Figure 2-4
Mayo Clinic's general etanercept page.

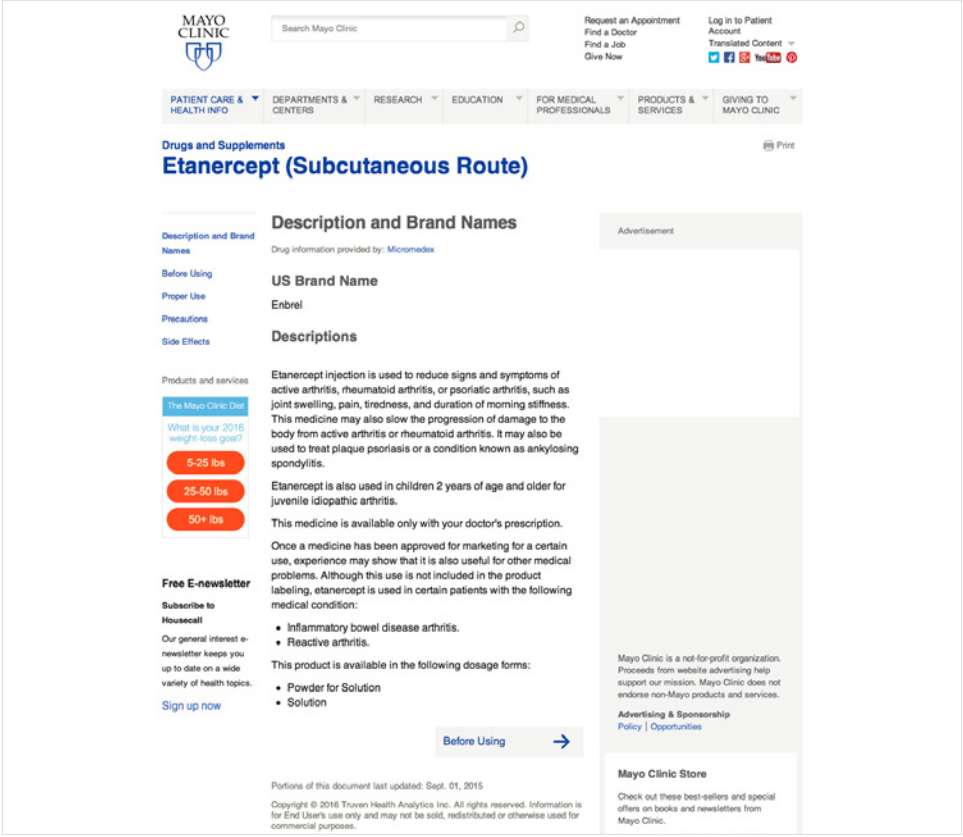


Figure 2-5
RxList's general etanercept information page.

Mayo Clinic

mayoclinic.com

Mayo Clinic pages on medications, such as their etanercept page in Figure 2-4, are similar to those of WebMD and Drugs.com. This Web site also licenses content from a drug information provider, Micromedex. Similar to Drugs.com, the bulk of information is about precautions and contraindications. Limited side effect information is available, again with no sense of frequency, magnitude, or duration. On Mayo Clinic's RA treatment overview page (see Figure 2-4), there is a straightforward description of lifestyle, medical, and surgical treatments. As far as medication information is concerned, only the briefest overview of RA drugs is presented. The content is entirely textual.

RxList

rxlist.com

Another resource frequently found on the first page of general medication search results is RxList, which is part of the "WebMD Network" of Web sites. That network includes other top search results MedScape and MedicineNet. Figure 2-5 shows a typical drug information page on RxList. In addition to ads and a large number of links and pointers to unrelated content, the page is effectively a modified version of the SPL supplied by the manufacturer when the drug. It has information entirely non-relevant to patients (such as the apparent molecular weight of the etanercept protein) along with extensive textual information. In general, the text is long and filled with jargon. Practitioner-targeted information coexists with patient-targeted information in a potentially confusing way; there are also menu items for both "consumer" and "patient," which repeat and rephrase much of the information found elsewhere, with slightly different organization. Finding specific information, such as whether the drug is safe given a particular contraindication or compatible with a lifestyle preference, requires reading much of this content.

Figure 2-6
A side effect table from the product SPL on RxList.

REACTION	PLACEBO CONTROLLED[A] (STUDIES I, II, AND A PHASE 2 STUDY)		ACTIVE CONTROLLED[B] (STUDY III)	
	PLACEBO (N = 152)	ENBREL[C] (N = 349)	MTX (N = 217)	ENBREL[C] (N = 415)
	PERCENT OF PATIENTS		PERCENT OF PATIENTS	
Infection[d] (total)	39	50	86	81
Upper Respiratory Infections[e]	30	38	70	65
Non-upper Respiratory Infections	15	21	59	54
Injection Site Reactions	11	37	18	43
Diarrhea	9	8	16	16
Rash	2	3	19	13
Pruritus	1	2	5	5
Pyrexia	-	3	4	2
Urticaria	1	-	4	2
Hypersensitivity	-	-	1	1

[a]Includes data from the 6-month study in which patients received concurrent MTX therapy in both arms.

[b]Study duration of 2 years.

[c]Any dose.

[d]Includes bacterial, viral and fungal infections.

[e]Most frequent Upper Respiratory Infections were upper respiratory tract infection, sinusitis and influenza.

Because this resource essentially republishes the SPL, if it contains quantitative data about efficacy or adverse reactions RxList features these data. However, they are represented exactly as the manufacturer supplied them, not redesigned to be more understandable by the general public. The adverse reactions table in Figure 2-6 is typical of such data. It includes terms like *pruritus, pyrexia,* and

urticaria, which are almost certainly unfamiliar to an ordinary person seeking information about RA medications. It is also likely that the data here are difficult for someone with lower health literacy to make sense of. For instance, multiple frequencies are reported for the same side effect. Careful reading of footnotes is necessary to understand that the studies were of different configuration and duration. This is precisely the kind of data that needs translation to be part of a data-driven health resource for the general public. It is worth noting that even here, the tabular data are *not* encoded in machine-readable form, so such translation is not possible from this representation of the data. Also, data like these—supplied by manufacturers—are not systematically produced, and may be less reliable to due bias, as discussed in the previous chapter.

MedicineNet

medicinenet.com

Yet another top search result resource is MedicineNet, also part of the WebMD family of Web sites. Its etanercept page (see Figure 2-7) is representative, and promises to "Bring Doctors' Knowledge to You." A superficial gloss of its design reveals choices remarkably similar to WebMD. Content about etanercept itself is textual. Only a fraction of the available screen space is even devoted to that content. Much more prominent and eye-catching are ads, "suggested reading" on other topics, and pictures captioned with phrases like "Don't Wreck Your Teeth."

PatientsLikeMe

patientslikeme.com

Unlike the most common online health information resources, PatientsLikeMe is not premised on "traditional" medical evidence (scholarly literature) but instead on real-world experience. Their medication page centers on data reported by its users—patients and their caregivers—for dimensions as varied as perceived effectiveness, side effects, burdensomeness, cost, and treatment regimens, among others (PatientsLikeMe, n.d.). The data are collected and encoded in a machine-readable way, enabling data-driven UI and visualizations. Thus PatientsLikeMe is an unusual resource because it is both data-driven, and based on a non-scholarly source of medical evidence.

General drug information (see Figure 2-8) is textual and licensed from a content provider, Cerner Multum. It is nearly identical to the content on Drugs.com and other online resources. The lack of UI parity—data-driven UI for self-reported information, and plain paragraphs of text for "official" information—is unfortunate, because it makes comparing across datasets difficult. For example, PatientsLikeMe has graphical presentation of patient-reported perceived medication efficacy, but no similar interface for data from medical literature. Indeed, efficacy would be reported using different measures, and therefore difficult to harmonize. But the point is that it is not even possible without accessible machine-readable research data to complement PatientsLikeMe's.

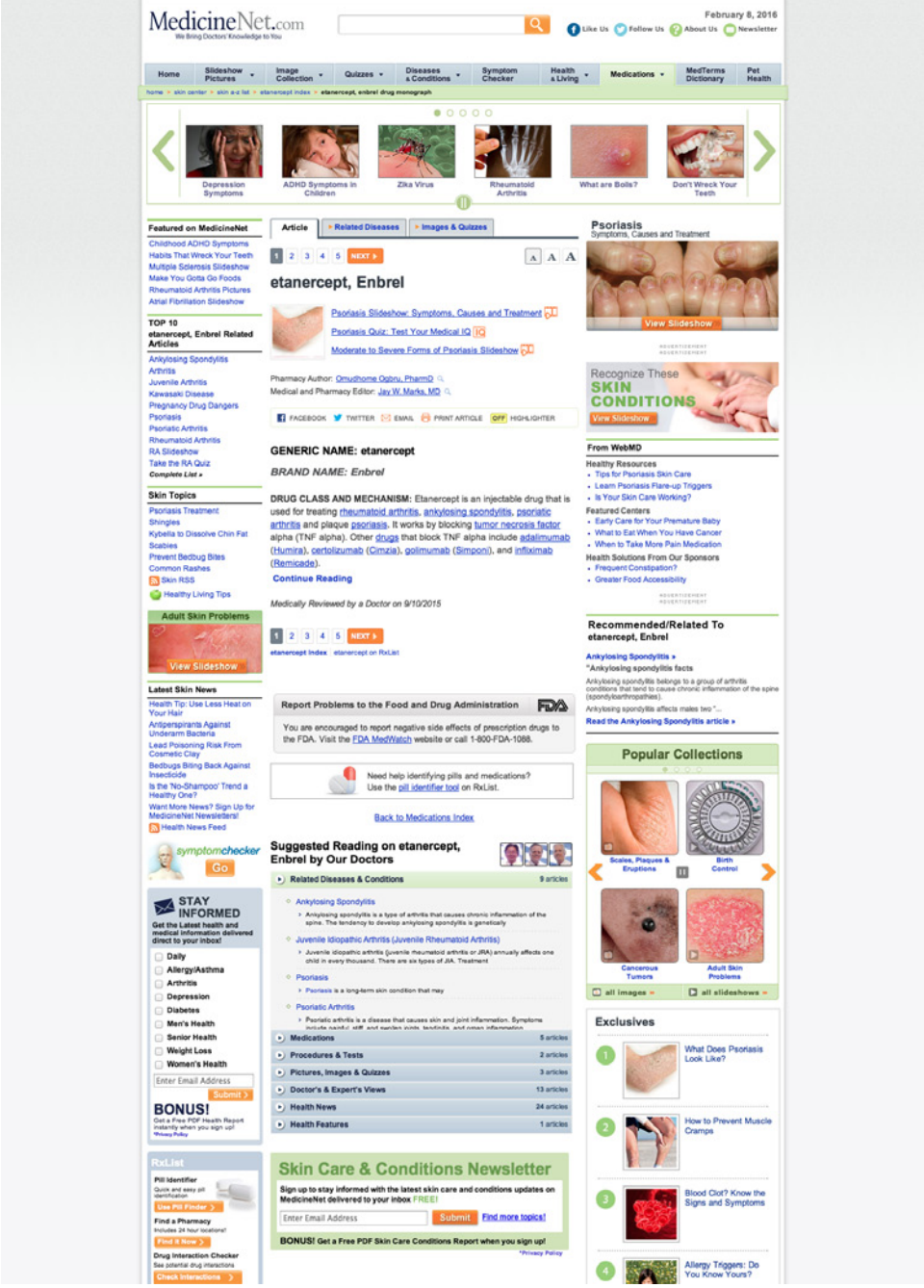


Figure 2-7
MedicineNet's etanercept page.

PatientsLikeMe has a page for RA (see Figure 2-9). As of mid-February 2016, it had 9406 users who reported having RA (89% female), including 5675 who claimed to be "diagnosed," although it is not clear how many had provided reports on symptoms and medications. However, it is a guide to symptomatology and relative perceptions of treatment effectiveness for those specific symptoms. The most common symptoms, ordered by "how bad" they are perceived to be, are stiffness in morning, joint pain, fatigue, pain, insomnia, depressed mood, and anxious mood. Patients report taking specific treatments to deal with those symptoms; unsurprisingly, none of them are csDMARDs or bDMARDs. However, people do report taking a diverse

range of medications–for pain, opioids like tramadol and hydrocodone/acetaminophen, and gabapentin, which is used to treat neuropathic pain; for anxiety, benzodiazepines, and even amphetamines (i.e. Adderall) for fatigue. This emphasizes the polypharmacy that many patients experience with when living with RA.

Under the heading "Compare treatments taken by people with Rheumatoid Arthritis (RA)," DMARDs *are* compared, along with glucocorticoids, pain relieving medications, and others. Although factors that would ordinarily be controlled in a clinical trial are absent (e.g. whether patients were taking other medications), it is possible to get a relative sense of the perceived efficacy and side effect burden of commonly used medications. For example, methotrexate, hydroxychloroquine, and sulfasalazine seem to be perceived as similar in effectiveness, with methotrexate and sulfasalazine apparently more burdensome in terms of side effects, especially nausea, which may be an especially concerning side effect for many patients (L. Fraenkel et al., 2004). Etanercept and adalimumab appear to be perceived as more effective, and slightly less burdensome in terms of side effects. Prednisone–a powerful anti-inflammatory agent–is perceived as much more effective, and surprisingly with a less severe side effect burden. For each treatment (for example, the etanercept page in Figure 2-10) PatientsLikeMe has a similar data-driven UI.

Because PatientsLikeMe does not have a publicly available API, it is not possible to digitally incorporate their machine-readable patient-reported data with the prototype, which would be an interesting complement to evidence from the scholarly literature.

Figure 2-8
PatientsLikeMe's etanercept drug information.

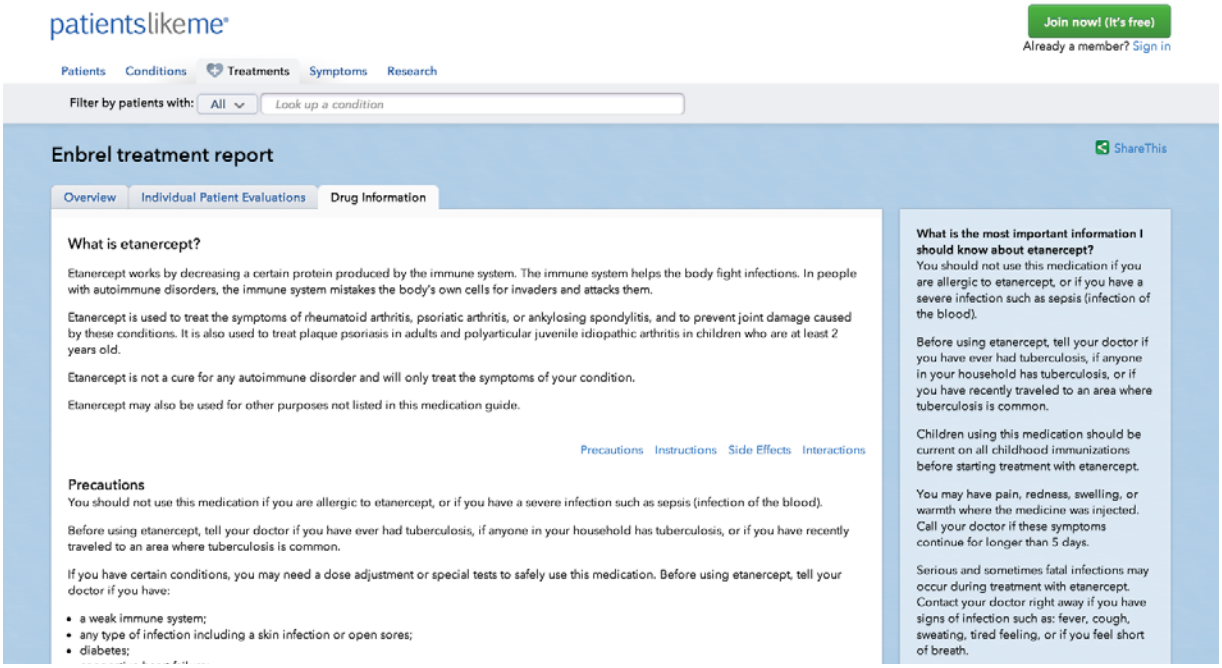


Figure 2-9
PatientsLikeMe's RA page. Patients' experiences with different medications can be directly compared with these visualizations.

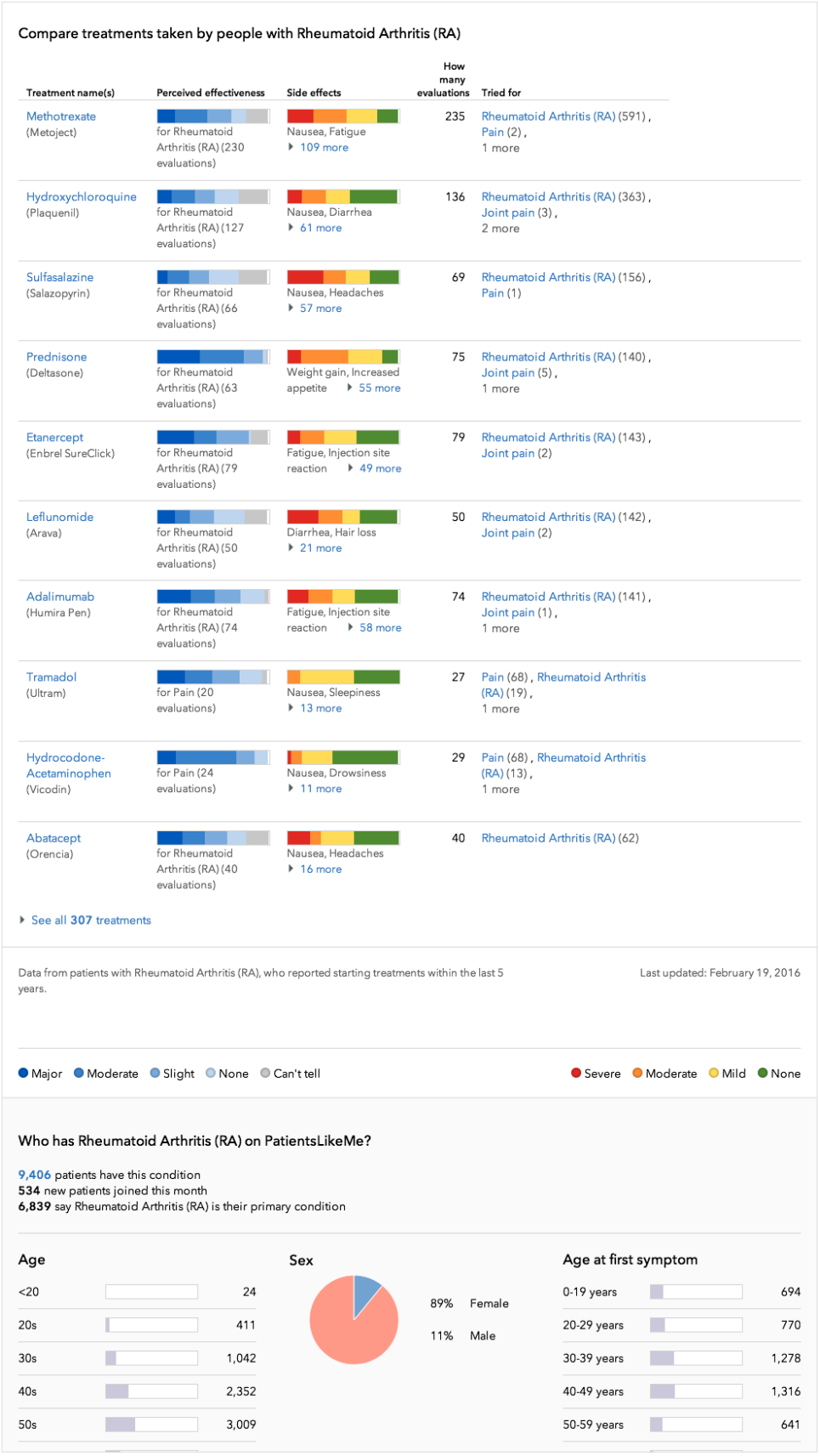
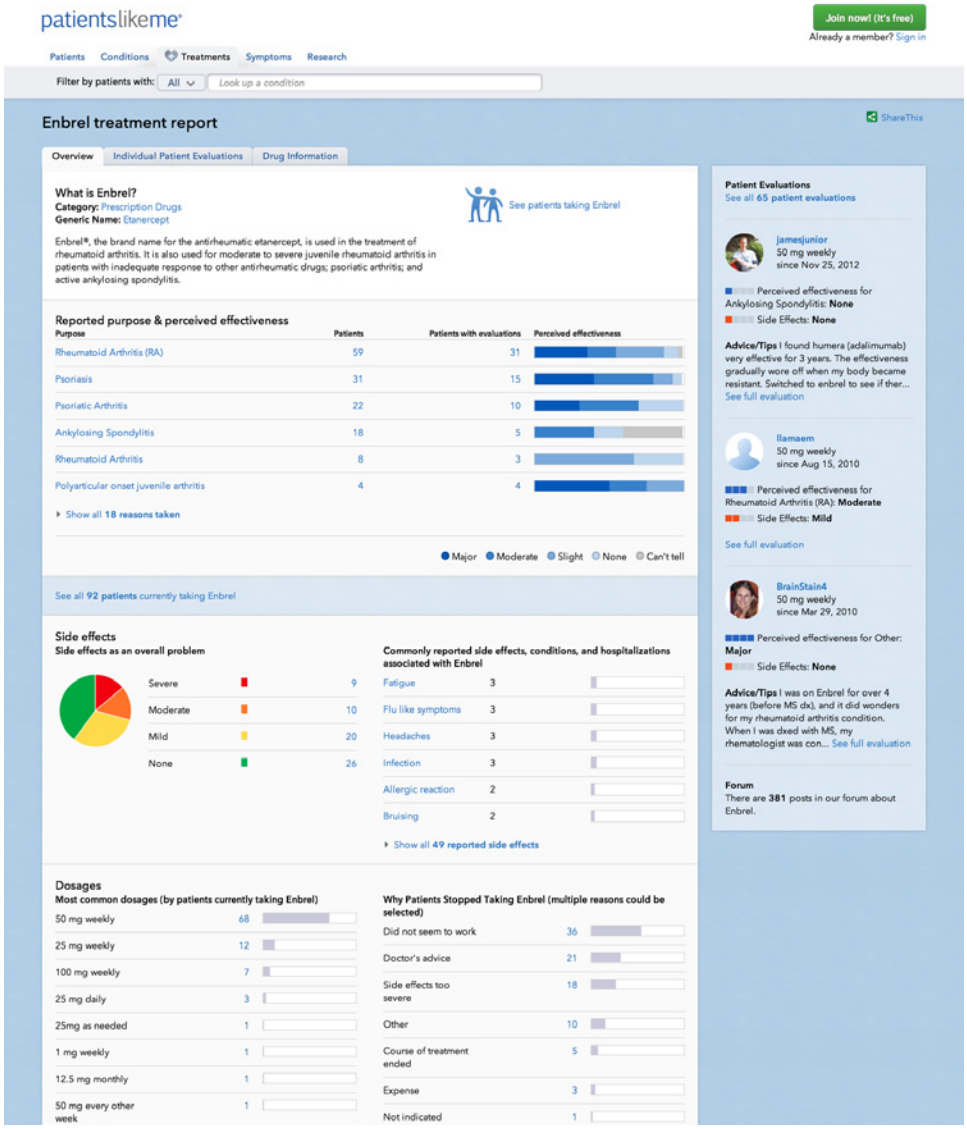


Figure 2-10
PatientsLikeMe's etanercept patient experience page.



CureTogether

curetogether.com

CureTogether is a Web site that collects patient experience data on therapies, and presents them in slightly interactive (sortable) visualizations (see Figure 2-11 for their page about treatments for RA). Their home page says that CureTogether is "the smarter way to find the best treatments," and that people can "get access to millions of ratings comparing the real-world performance of treatments across 637 health conditions." Like PatientsLikeMe, this service focuses on patient-reported data. An enormous variety of treatments are presented, from medicines, surgical procedures, and occupational therapy to dietary adjustments, deep breathing, and even hypnosis. Since each treatment has a certain modality and perspective on how it might improve living with RA (e.g. psychological or self-perception, disease modification, symptom relief,

improving general health, etc.) it is uncertain how to compare treatments to one another. For example, what does it mean that both “exercise” and “heat” have similar effectiveness profiles? On this resource, prednisone is perceived to be the most effective treatment–but again, that may be unsurprising considering its potency, and is potentially a misleading notion absent of the context that it is a potentially harmful medication when used for a long period of time at high doses.

CureTogether has no API, so its patient-reported data cannot be harmonized with similar data from a site like PatientsLikeMe, nor can it be included in the prototype alongside data from peer-reviewed medical literature.

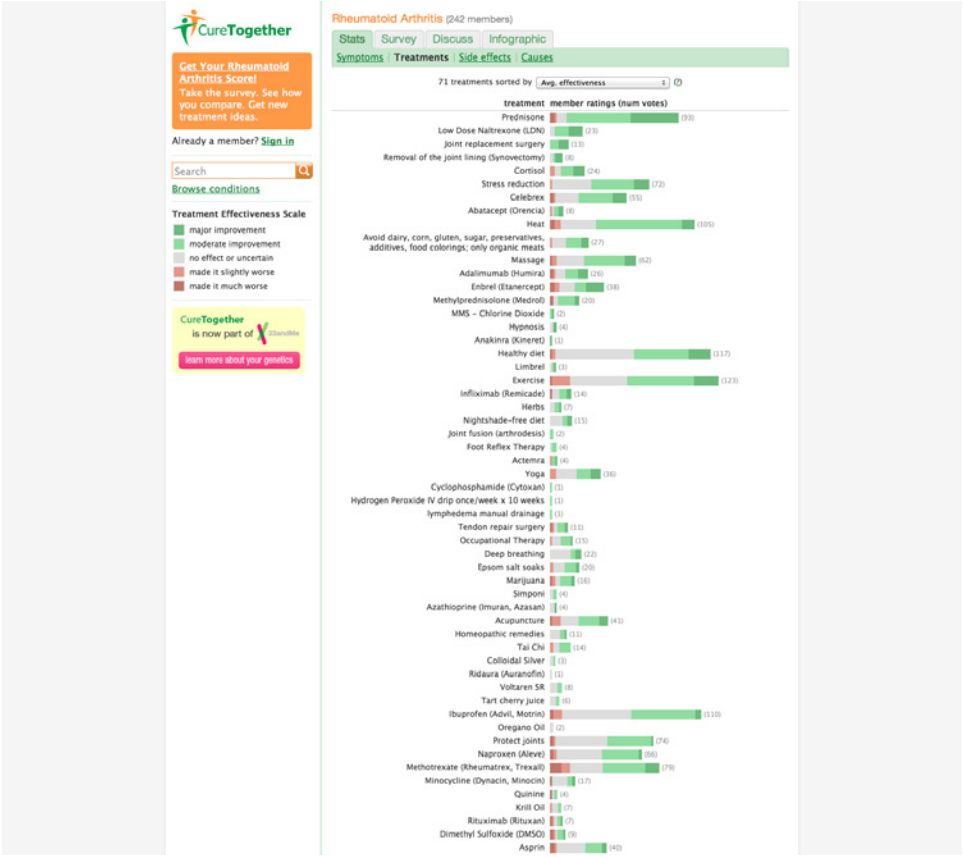


Figure 2-11
CureTogether’s RA treatment comparison page.

Healthtalk

healthtalk.org

Healthtalk.org (Healthtalk) is a Web site produced by a group at Oxford University that conducts qualitative research (see Figure 2-12 for its page about DMARDs). It blends brief textual summaries of up to date medical knowledge with extensive summaries, videos, and transcripts of interviews with people with RA about their experiences with everything from diagnosis to family impacts to treatments to attitude and psychosocial effects of living with RA. Although the video interviews are anecdotal, they are intended to be

representative of the “full range of experiences that might be connected with a health condition” (See About healthtalk.org on healthtalk.org, n.d.) presumably among the British population from which its sample of patients was drawn.

This resource is unique in that it has an information set based on qualitative research that is rigorously collected and reviewed by experts, as opposed to a freewheeling discussion or set of posts from patients. A patient experience resource with unrestricted membership might suffer from a kind of selection bias if only a particular subgroup of the population at large were represented; if the factors that made those members unique were not explicitly collected and communicated, it would not necessarily be easy to generalize information to the general public. For instance, because PatientsLikeMe indicates that about 89% of its RA members are female, it is not necessarily fair to say that

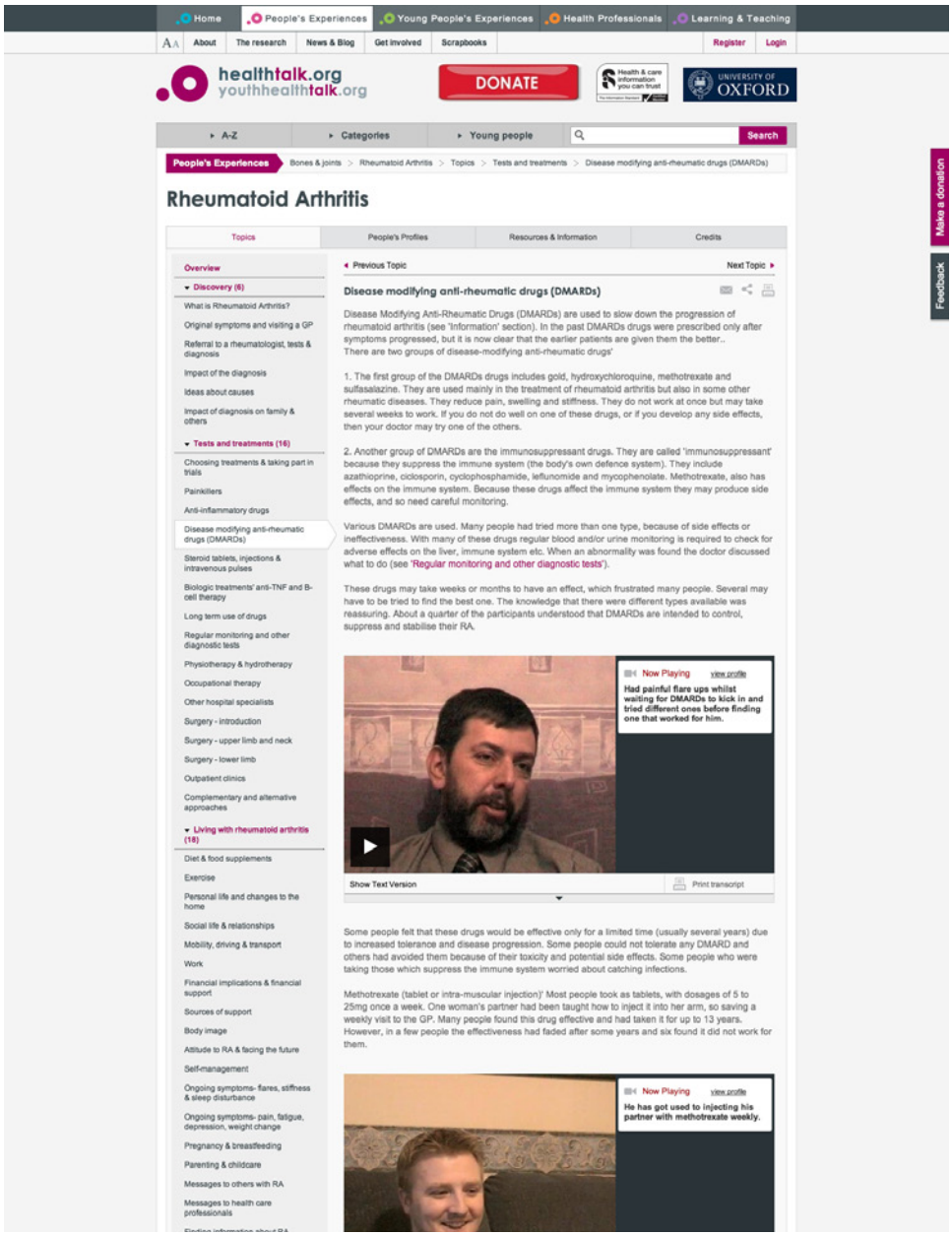


Figure 2-12
Healthtalk's DMARD page.

its “findings” are typical of the male population, or population at large. While Healthtalk has its own bias in terms of the population that was interviewed, its creators do apply greater rigor in order to build a representative resource.

As is the case with many decision aids, the information in Healthtalk’s pages are in effect fixed to a point in time, when the interviews were conducted and the pages created. Because much of its information is general in nature (e.g. perceptions of body image while living with RA) that may not be an issue. However, when it comes to information about diet, medications, treatment regimens, and so forth, it may be advantageous for a resource like this to accommodate new, machine-readable evidence.

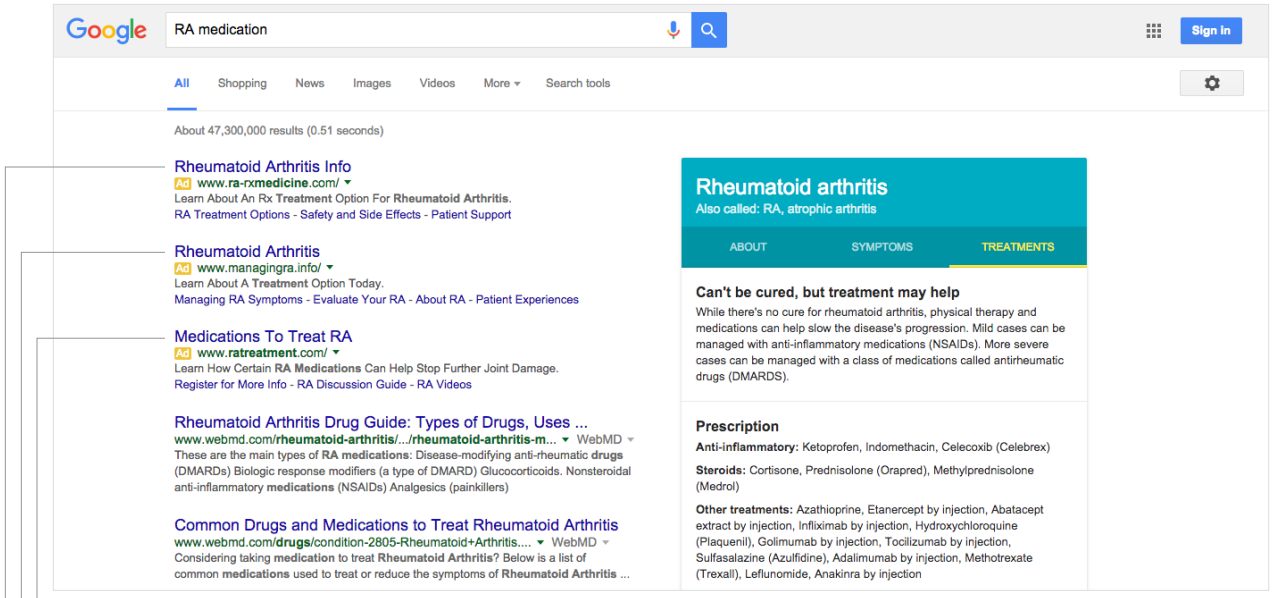


Figure 2-13
A Google search results page with masked manufacturer results.

Points to a product Web site for Humira, the brand of adalimumab manufactured by AbbVie.

Points to a product Web site for Orencia, the brand of abatacept manufactured by Bristol-Myers Squibb.

Points to a product Web site for Xeljanz, the brand of tofacitinib manufactured by Pfizer.

Healthtalk’s information (such as textual summaries, transcripts, and video interviews) are themselves not machine readable data, which limits the ways in which they can be easily automatically incorporated with other data. Even qualitative information–about attitudes, preferences, and so on–*can* be systematically encoded in machine-readable formats, but it is not obvious just how that should be done. For example, without sufficient metadata for the video interviews (descriptive, discrete data *about* the video, such as the topics covered, the patient demographic data, treatments the patient was on, etc.) it is not easy to “pull” the data into another resource such as the prototype, to appear alongside other evidence. Regardless, Healthtalk does not have an API, so there is no easy machine-machine interface for bringing its information into another resource.

Product Web sites

Alongside the resources described so far, medication Web sites published by their manufacturers are commonly found on the first page of Google search results. Sometimes they are “organically” ranked highly, alongside resources like WebMD, but often they are marked as paid (advertising) links. Drugs that

are recognized by their brand names and protected by patent dominate such results. Searches for RA medication are liable to return a search results page with paid links to manufacturer sites. The links are often disguised in that they do not mention that the name of the medication, or the manufacturer, directly on the search result page. For example, in Figure 2-13, the first paid link points to the Xeljanz product Web site, an oral DMARD marketed by Pfizer. The link title is "Rheumatoid Arthritis Info," and the URL (Web site address) is "www.ra-rxmedicine.com" with a snippet (description) that says "Learn About An Rx Treatment Option For Rheumatoid Arthritis." As of February 2016, according to the Google AdWords Keyword Planner tool, the "suggested bid" price for a single click on a paid search result for the search query *RA medication*–which had an estimated 880 searches per month in the U.S.–was $28.84.

The product Web sites that these links point to may be more interactive, thorough, and even data-driven than other Web sites. The Xeljanz Web site (see figure 2-14), for example, has an interactive and visual explanation of different classes of medications used to treat RA, how they may be used together, and that describes when biologic and alternative DMARDs (like Xeljanz) are generally considered (i.e. if methotrexate does not work well). It features data from (presumably favorable) clinical trial results–data absent from resources like WebMD. It would be worth studying whether people understand that these Web sites are published by manufacturers, what their perception is of the potential bias or influence of those manufacturers is, whether the data they see are convincing, and whether they consider these resources trustworthy.

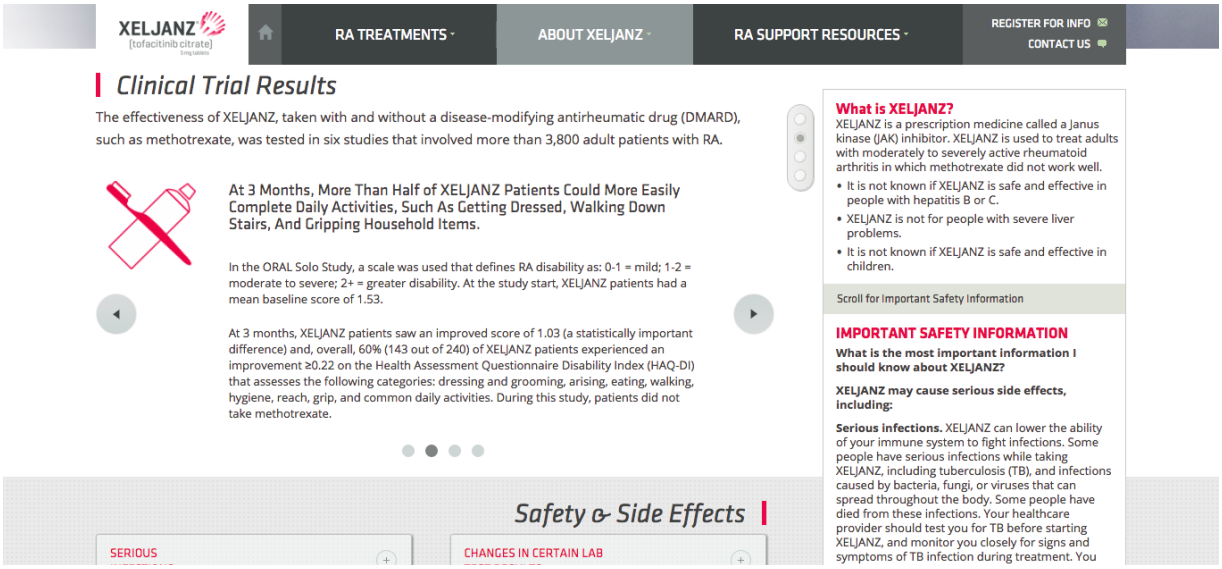


Figure 2-14
A page from the Xeljanz product Web site.

Data-driven non-medical online applications

Outside the health and medical domains online, data-driven resources commonly populate the first few search results. Web sites and apps for sports, shopping, travel, weather, news and politics, mapping and navigation, and day-to-day business often feature rich multivariate data. Often, these resources incorporate historical and real-time data; they can therefore represent the current state of affairs, rather than a static, fixed perspective of some point in the past. Certainly, many of these resources are built with data that are not subject to the same review and analytical rigor that medical data are. Nevertheless, there are lessons to be learned from their UIs and how they help non-experts make meaning from their underlying data. Deliberate design decisions again shape information presentation, and the ways in which people can and cannot interact with it.

Fundamentally, the best such resources employ data that can directly answer user questions. For example, questions of cost can be answered with accurate price data and clear, unambiguous presentation. Many non-health consumer resources are thus at an advantage compared to potential online health counterparts, for which straightforward data are often unavailable. Still, even directly applicable data need translation, recombination, and appropriate UIs to visualize, restrict, and clarify meaning. An exaggerated example illustrates the central problem.

Before the advent of smartphone apps and GPS for transit navigation, how would one make sense of the data which underlie such apps?

- **Mapping data.** The raw data describe the coordinates—longitude and latitude—and shape of roads, along with tables recording the speed limit on road segments.
- **Traffic data.** Raw historical traffic data might be recorded as tables describing the average speed of vehicular traffic on road segments. Just a few short years ago, real-time data might be found only ephemerally in live radio traffic reports.
- **Transit route data.** Such raw data would record the direction that routes travel on road segments, along with the location of stops.
- **Transit schedule data.** Most likely, such data would be recorded as scheduled departure times at certain stops, or at certain coordinates.

To make sense of these data, each source must first be translated into an initially useful form. If the data were presented in tables on paper—or worse, each type individually as a narrative—it would take a great deal of work to make meaning from them. Imagine trying to piece together a bus route by reading a table describing the road segments it follows, or a description of its route: "… continues northeastward on Folsom St. for 760 feet, then turns left onto 9th St. Continues northwestward on 9th street for 1930 feet, then merges onto…" Cartography neatly handles such road and transit route data in the form of familiar maps. Historical traffic speed data could be presented on a separate map, or in a table. The transit schedule would most likely be found in a table.

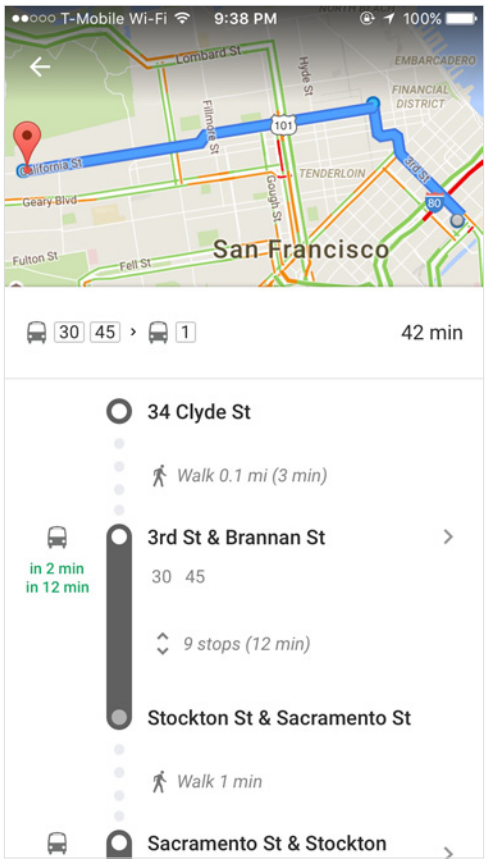


Figure 2-15
Integrated navigation data on Google Maps.

Real-time estimates, if available, might be available by phone, just as real-time traffic reports could be tuned into on the radio.

Given a simple problem—which buses should I take to get to my destination, how long will it *truly* take me to get there, and when must I leave to catch the first bus?—there are a couple of approaches. One is to gather the various sources and assemble the information. In this case, that is not an exceedingly difficult task but would probably require a few minutes, cross-referencing, likely and some double-checking. Consultation of the real-time broadcast traffic report—for a more accurate estimation of travel time—would not be possible on-demand, so depending on when the report is broadcast one might have to wait awhile for the information. In the end, the gist of the answer would emerge. Another approach would be to consult an expert—to telephone the transit agency's information service, or ask a knowledgeable friend—who could express the navigation plan and estimated arrival time narratively. (And one could take down notes of that narrative, to refer to in transit.) Individual knowledge of the roads, traffic and travel times, routes, and schedules would factor into the ultimate analysis, too.

It almost goes without saying with machine-readable data and a clever UI, a computer can do much of this work on behalf of the traveler—still allowing space for their individual knowledge and judgement. Instead of looking at a map on which the routes in question are not highlighted or differentiated, the information can be tailored. A map can be produced showing just the segments of the routes the traveler needs to take, with annotations showing real-time departure information and guidance that helps one know what to expect from the trip (see Figure 2-15 for such an example from the Google Maps application for iPhone).

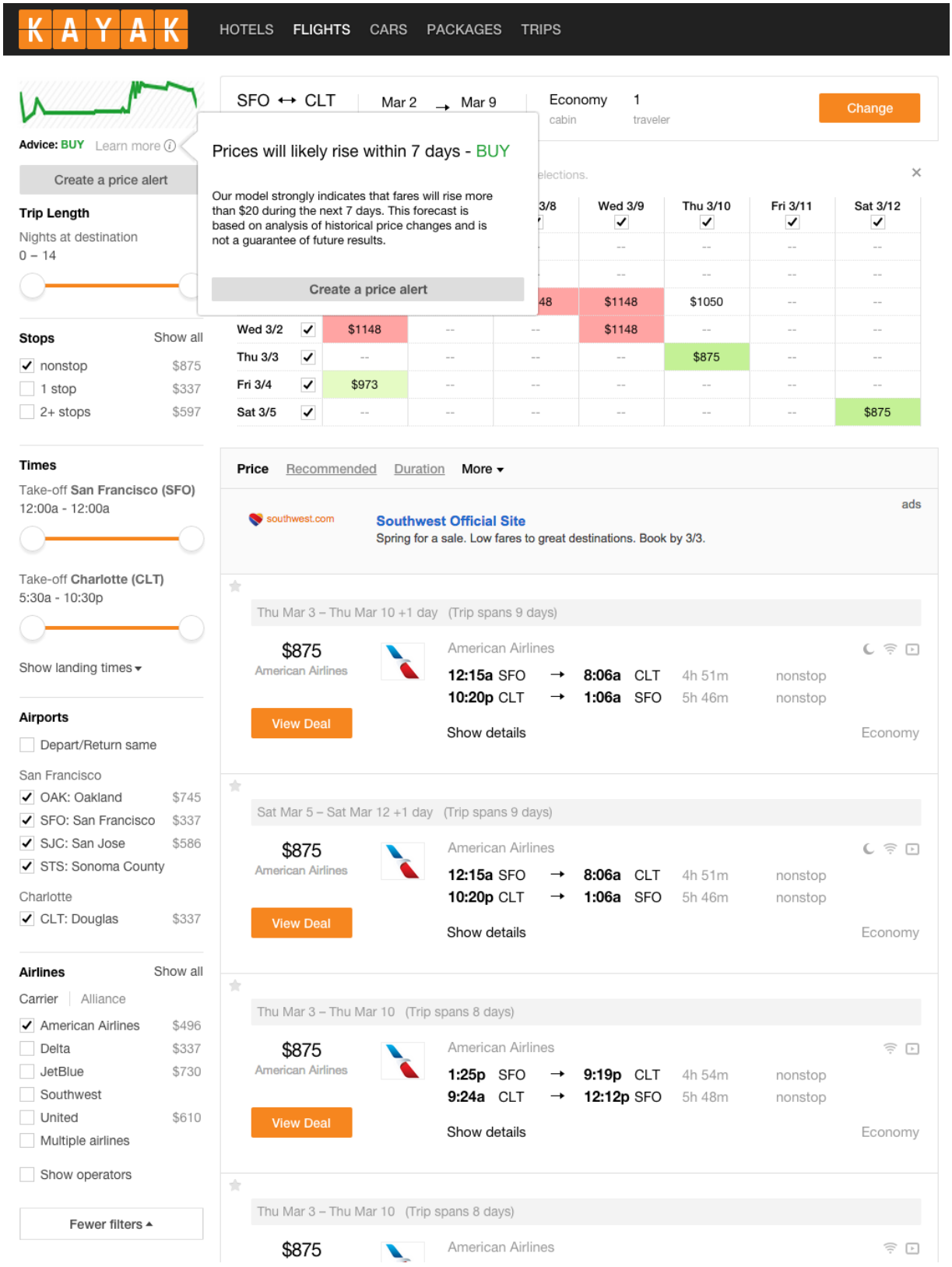
While this example is admittedly difficult to transmute to medicine, it represents an essential point: Non-experts *can* work with "data" if they are encoded for machine-readability, and brought together in appropriate UIs. Because an inventory of such products is beyond the scope of this project, a few examples of data-driven resources are subsequently briefly reviewed to discuss their design and applicability.

Kayak

kayak.com

Kayak is a Web-based travel booking tool—for flights, hotels, rental cars, and vacation packages. Its users interact with vast quantities of data to filter and narrow options and support travel decisions. The rest of this brief discussion focuses on its ordinary flight search. It features a faceted search UI, which means that it provides user interface controls to filter or limit flight results, for example, along dimensions like airline, number of stops, price, flight

Figure 2-16
Kayak flight search results. Purchasing advice accompanies faceted search and multiple data visualizations to filter and make sense of available flights.



duration, origin and destination airports, class of service, and so on. Results are presented in a more or less textual format, with greater visual weight given to certain pieces of information that are presumably more valuable for decision making. Price is the largest data element in a search result, and flight times boldest; when a certain filter is applied, results update immediately, hovering the cursor over a filter highlights data elements related to that filter in orange. Kayak features sophisticated analysis tools, such as a 7-day-by-7-day matrix that highlights relatively inexpensive and relatively expensive flights. It also automatically draws attention to factors that may impact a traveler's decision but that are not obvious. Common situations include when the departing and returning airports are different, when a layover is especially short, when an aircraft is a turboprop where a turbojet may have been expected, and when a flight spans the international date line. Although each search result incorporates more than a dozen individual data elements, low-contrast color helps less critical data recede. Progressive disclosure is employed to reveal even more information that is usually irrelevant to the first pass of decision making, such as the flight number and aircraft type. A "purchase advice" visualization shows a historical price sparkline (seen in the top left of Figure 2-16), accompanied by an unambiguous recommendation—"BUY" if Kayak's forecast of prices suggests that the price will rise in the coming days.

Kayak facilitates interaction with data that are dissimilar from most of the kinds of outcome evidence someone might encounter in the course of researching medications. However, it does demonstrate techniques for manipulating and filtering data, and presentation of the data in ways that can supply meaningful information to a decision—in ways that are presumably accessible to the general public, which Kayak serves. Notably absent is experiential knowledge (e.g. challenges getting to and from airports, whether one will have to pass through customs for international transfers, or reviews of flights or aircraft) that could also aid decisions.

Hipmunk

hipmunk.com

Hipmunk, like Kayak, is a travel booking Web site. Its flight search UI (see Figure 2-17) is also faceted, but differs in that it relies more heavily on visualization. Individual flights are visualized on a timeline, so that the number of segments, layovers, and relative duration can be compared by their length on screen. Additional details—like aircraft type, on-board amenities, and so forth—are revealed only upon request. In this way, Hipmunk's designers have made deliberate decisions to focus on information that they either believe or have found are important to travelers. Price is, of course, prominent, and relatively inexpensive flights highlighted in green. Like typical search result UIs, Hipmunk's can be sorted by facets like price, duration, and takeoff time. However, it also introduces a novel sorting algorithm—called *agony*—which is the default method by which flights are sorted, suggesting that the first result is the least agony-inducing. On its frequently asked questions page, Hipmunk says "we know that price isn't the only factor that goes into purchasing a flight," and that its notion of agony is "primarily a combination of price, flight duration, and number of stopovers" (Hipmunk, n.d.). Here, Hipmunk demonstrates a kind of synthesis of multiple pieces of data into a gist which attempts to map

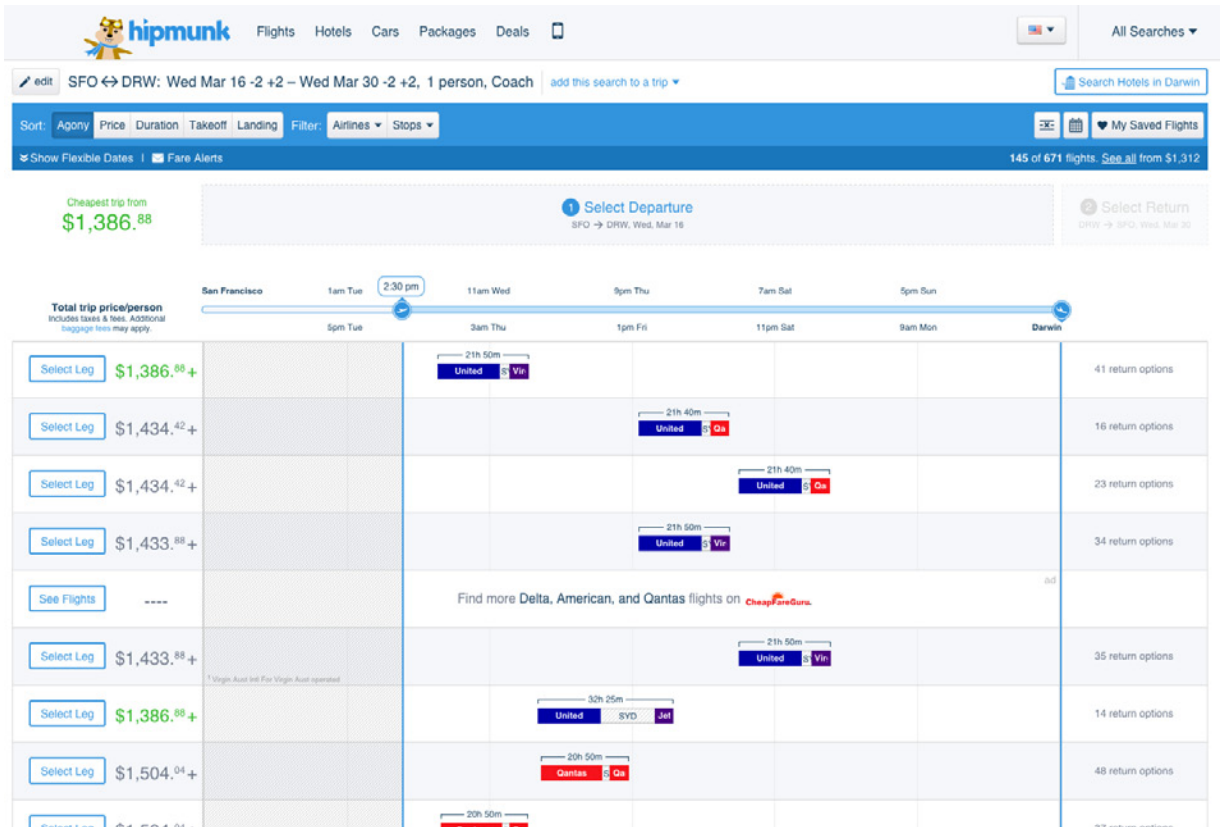


Figure 2-17
Hipmunk flight search results.

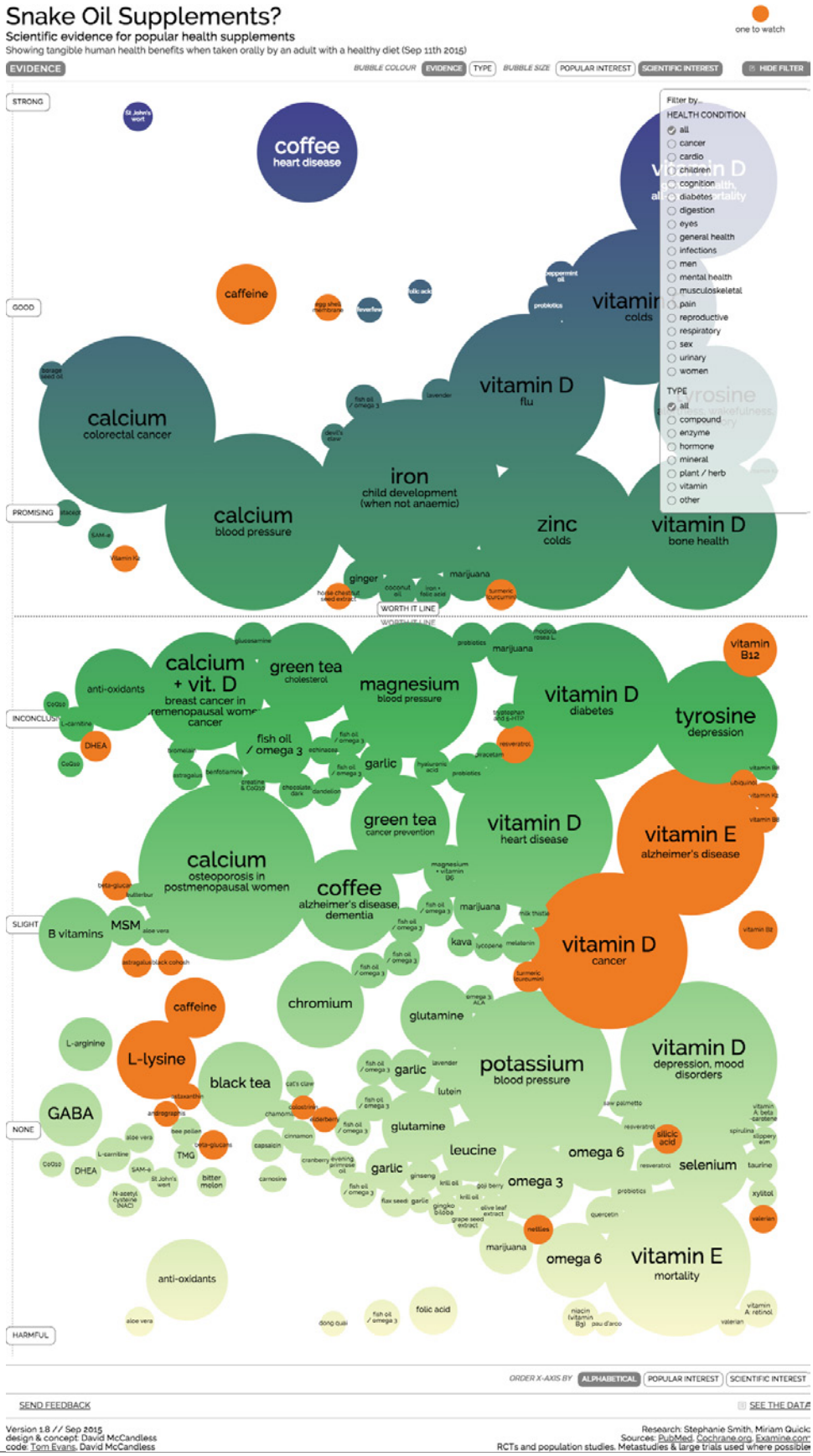
to the mental model of its traveler users. It features a similar system for hotels, called *ecstasy*. Both concepts are examples of translation from jargon (or raw data) to a form that is intended to help an audience make meaning with the data.

Snake Oil Supplements

www.informationisbeautiful.net/visualizations/snake-oil-supplements

Snake Oil Supplements (SOS) is an interactive data visualization (see Figure 2-18 for a static snapshot) that summarizes evidence about 191 pairs of a non-pharmaceutical supplement or remedy and a specific health concern. For example, it has information about St. John's wort for the treatment of depression, based on data from a systematic review published by the Cochrane Collaboration. Each supplement-health concern pair is represented by a "bubble" on the visualization. The bubble is roughly positioned vertically according to the strength of evidence for tangible health benefits, according to a score marked by the visualization authors, who manually reviewed the evidence. The scale has 7 discrete categories: Harmful, none (no evidence of health benefit), slight, inconclusive, promising, good, and strong. The bubble is sized according to either relative popular interest (measured by estimated Google search query volume for a query related to the supplement-health concern pair) or scientific interest (measured by the number of citations on Google Scholar from 2000 to 2012 for a query related to the supplement-health concern pair). Like the prototype, SOS uses a Google Spreadsheet as its "back

Figure 2-18
Snake Oil Supplements.



end" or data store. Anyone can look at the visualization, and the provenance of evidence that populates it. For each pair, one can find links to the literature and quotes or abstracts that elaborate on the summarization of the evidence. Like Kayak and Hipmunk SOS features a faceted filtering UI to restrict results.

Effective translation is at work in more than one way in this visualization. First, the evidence is summarized in a way that maps onto a simple question: "Is this supplement helpful for this health condition?" Summarizing evidence in this way is complicated and eliminates almost all nuance from its final presentation. However, one can look at this and see at a glance that there is a good deal of scientific interest in vitamin D supplementation for general health and to reduce all-cause mortality, for which strong evidence for a benefit (according to the SOS authors' review) has been found. St. John's wort for depression, by comparison, has seen relatively little scientific interest but is backed by similarly "strong" evidence. Second, a line called the "worth it line" bisects the visualization. Supplement-health concern pairs with at least "promising" evidence are above the line, suggesting that they are more likely to be good for health and worth taking. Below that line, the evidence suggests that it may not be worth it, either because there is conflicting data or because the supplement has been found ineffective or harmful—regardless of the level of interest in it. This is a novel and straightforward way of mapping the data onto an ultimate, very human decision about whether to take a supplement or not.

The visualization raises a host of questions about evidence selection, summarization, and so on. Because as a whole it is a system so similar to the prototype, most of those questions are dealt with in detail in Chapter 3. How this product differs is that it is not a generalized or systematized platform for developing further visualizations based on the same data. Instead, it is a bespoke visualization built on an expandable "evidence base" with summarizations encoded in a spreadsheet. Because the summarizations—of "promising" evidence of health benefits—are encoded in the spreadsheet instead of individual findings—of specific depression relief outcomes for St. John's wort, for instance—only those encoded summarizations can be compared. That is suitable for a visualization that is intended to communicate a high-level notion about dozens of disparate supplements, but not as suitable for a platform on which more detailed visualizations might be built to explore specific findings in the source evidence.

How Likely Is It That Birth Control Could Let You Down? (The New York Times)

www.nytimes.com/interactive/2014/09/14/sunday-review/unplanned-pregnancies.html

Another data visualization project produced by a news organization, this birth control visualization by The New York Times (see Figure 2-19) is more akin to the kind of resource that might one day be available for general medical inquiries. It is outcome-oriented, in that it communicates a potential outcome (unplanned pregnancy or contraceptive failure) given certain conditions (typical or perfect use of the contraceptive method). It uses multiple visualizations and annotations, along with a distinctly interactive UI (hovering the cursor over visualizations reveals data that facilitates understanding of the charts). The repeated and consistent presentation for each contraceptive

method promotes comparison from one method to another. It is data-rich, and probably requires significant health or data literacy for full understanding. Nevertheless, it is an example of a resource that might inspire outcome-oriented resources in other domains–for example, it is relatively easy to imagine how the same scaffolding could be applied to estimates of effect about RA medications given specific conditions of medication adherence, dose, or other factors.

Figure 2-19
New York Times birth control efficacy visualization.

Related work in medicine

Evidence-based medicine

In the 1990s, alongside the emergence of the Internet for the general public, evidence-based medicine (EBM) began ascent. It is a widely adopted model of medical decision making that prioritizes high-quality medical evidence over clinical experience or pathophysiologic rationale (Davidoff, Haynes, Sackett, & Smith, 1995; Eddy, 2005; Guyatt et al., 1992). High-quality evidence is that which is *ideally* free of bias, statistically sound, and applicable to the decision at hand. Well-conducted and transparent randomized clinical trials, systematic reviews or meta-analyses of large bodies of evidence, and clinically applicable epidemiological studies might all be sources of such evidence. With EBM, clinical decisions can be made with current best evidence made visible to the patient, instead of strictly by intuition. EBM *is* practical in that it can be practiced today, however it depends on extant evidence for the decision at hand. When there is no good evidence available to help answer a clinical question, and yet a decision must still be made, clinical intuition and pathophysiology may be all that a practitioner can rely on. An illustration depicts why evidence-based practice can be beneficial:

> I would like to retell a story that was told to me by Sir Iain Chalmers – one of the people who have worked so hard to make Archie Cochrane's vision a reality in the form of the Cochrane Collaboration. Chalmers told me how as a young doctor he bought a copy of Benjamin Spock's famous book Baby and Child Care. Spock was an American paediatrican and his book, first published in 1954, has sold 50 million copies in 39 languages and has been described as one of the most influential books of the 20th century. The young Dr. Chalmers marked the passage that advised mothers to put their babies to sleep on their tummies, advice he duly passed on to his patients. The rationale given by Spock was that babies put to sleep on their tummies would be at lower risk of inhaling vomit and choking, should they happen to vomit in the night. However, by the 1970s and 1980s evidence was accumulating that this, untested theory, was lethally bad advice. We know now that around 50,000 cot deaths worldwide were caused because of it. In fact it is much safer to sleep babies on their backs, a finding which completely reversed Spock's health care advice on the topic.
>
> (Barratt, 2008)

An increasing volume of evidence is produced every year (Davidoff et al., 1995). For it to be applied, it must be systematically reviewed and incorporated into resources that clinicians—and perhaps patients—can use. For many of the reasons cited in the earlier section "Factors that shape or limit access to medical evidence," much necessary evidence remains out of the reach of medical practitioners.

When high-quality evidence *is* available, it must still be considered in light of the individual needs of a patient. While a patient might "fit" the mold of one

in a population for which there is good evidence about an intervention, that person is still an individual—with individual capacities, needs, values, and preferences; a unique life.

Patient-centered care

The constellation of ideas that may be conceptualized as patient-centered care (or patient-centeredness) are those that promote care of the whole person, a positive therapeutic alliance between the patient and clinician, and often shared responsibility for decision making (Balint, 1969; see review in Holmström & Röing, 2010). The goal of patient-centeredness defined by the Institute of Medicine in *Crossing the Quality Chasm: A New Health System for the 21st Century*, is "to customize care to the specific needs and circumstances of each individual, that is, to modify the care to respond to the person, not the person to the care" (Institute of Medicine, 2001). These notions may be understood in contrast to care that excludes the patient's experience of illness, that is directed or guided by needs or impetus other than the patient's, or that is about treating a pathology or a disease. If the patient is seen as host to an illness, care is probably not patient-centered. Put another way, curing an illness at whatever cost is insufficient for care to be patient-centered. The patient's needs, irrespective of their illness, must factor into the practice of care.

A related but different notion is patient *empowerment*, which according to a review of the concept by Holmström and Röing (2010) deals more with patients' assumption of authority to enact health-related decisions, and with ensuring that people have the agency and freedom to educate themselves about and "own" such decisions. The empowered patient may not take control of medical decisions, but may feel capable of educating themselves and acting on their own behalf.

Achievement of patient-centeredness requires practices and health resources that acknowledge the person being treated, inform and educate, elicit needs and preferences, facilitate good working relationships and communication, provide physical and emotional comfort, and shape and deliver appropriate care. If someone wishes to be *more* informed and involved in medical decision making, ideally it should be so. If someone wishes to be *less* informed and *not* involved in medical decision making, ceding as much authority as possible to their clinician, ideally it should be so. In both cases, the patient is in charge. Similarly, a certain degree of patient empowerment depends on a medical-social willingness to recognize a patient's prerogative to take on and encourage a sense of self-efficacy in care responsibilities. Patient-centeredness and EBM could be seen as potentially conflicting (Barratt, 2008), but only if EBM is adopted in a dogmatic fashion, where a decision can only be made *according* to the best available evidence, instead of *considering* the best available evidence. Practices like shared decision making (SDM), aiming for goal concordance between physician and patient, and minimally disruptive medicine (MDM), along with tools like decision aids (DAs), and research oriented around patient-centered or patient-reported outcomes (PROs), are all part of the contemporary patient-centered context in which the prototype—and all online health information resources—are situated.

Shared decision making

Shared decision making (SDM) describes a model of medical decision making in which clinicians and patients collaborate on decisions (Charles et al., 1997; Stiggelbout et al., 2012). It is not a single method, but a way of practicing that involves give and take between providers and patients according to the individual values, needs, and preferences of both parties—and often additional participants, like caregivers and family members. Patients may "share" the decision by agreeing to vest authority with their clinician (Kon AA, 2010), but in effective SDM patients may be more likely to engage in a discussion about what may work best with their provider and more truly "share" the decision.

Because it is a patient for whom a medical decision—to perform a test, to start or stop a medication—has the greatest consequences, it seems obvious that they should be informed participants and ultimately decide what is right for them. But neither patients nor their providers can be expected to have all the answers, as is clear from the complex decision scenarios outlined in Chapter 1. Important medical decisions "must be made in complex, ambiguous clinical situations, in which clinical evidence is insufficient, goals and options are not clearly defined, and preferences are contextual, provisional, and conditional" (Epstein & Gramling, 2013). Though no single definition of SDM exists, it can be understood as the process by which these decisions can be made by the provider and patient sharing information, discussing the benefits, risks, and implications of intervention options (including doing nothing, if desired), and ultimately coming to a preferred option in line with the patient's individual needs. SDM is inherently a subjective, very much human endeavor that reflects the beliefs of every participant—that is, both clinicians and their patients. Some believe that persuading patients with evidence is an imperative (Shaw & Elger, 2013) especially where their beliefs may conflict with the state of scientific knowledge, while there remain significant challenges in bringing EBM and SDM together (Barratt, 2008). A good deal of the inquiry into SDM has to do with tools by which clinicians and their patients *can* use evidence to make informed decisions, including research on decision aids (more to follow below).

Inquiry into SDM—how patients and clinicians collaboratively make decisions, including by integrating medical evidence and tools like decision aids—has shown that various implementations improve the quality of clinical interactions from the patient perspective, although it does not necessarily produce better outcomes (LeBlanc A et al., 2015; Lin & Fagerlin, 2014; Shay & Lafata, 2015; Stacey et al., 2012). Regardless of the impact of SDM on health outcomes, because it is a present model of decision making, whatever resources (from WebMD to decision aids) patients use to inform themselves play a role in it.

Goal concordance

An issue key to collaborative medical decision making is goal concordance, or the degree of alignment between clinicians and patients on care objectives. It cannot always be assumed that these stakeholders' treatment goals are the same. While it may be clear to a physician that there is a single most effective course of treatment for the underlying disease, it may not be clear that the physician's patient or the patient's family agrees that it is the best option. Opposition could be due to overall treatment burden, cost, side effects,

unwillingness to change from a current therapy, lack of confidence in the treatment, or any number of other reasons. Although there is limited research on goal concordance, Heisler et al. (2003) found that in a sample of diabetes patients and their providers, there was considerable lack of alignment on goals—only 5% of patients and provider pairs independently agreeing on their top 3 treatment goals, with 19% having no overlap at all. Similar discordance was found for treatment strategies. However, they also found that patients who more strongly believed in the efficacy of their treatment were more likely to share goals with their provider; patients who agreed with their provider on the top treatment strategy had a greater sense of self-efficacy.

Greater goal concordance might play a role in effective SDM and greater patient-centered care. A natural question is whether it is plausible that a shared understanding about treatment options, built from a shared base of evidence, could contribute to alignment on treatment goals. In the case of RA treatment, could clear physician-to-patient communication about treatment strategies (e.g. recommendation of methotrexate as a first-line treatment, treat-to-target, use of multiple DMARDs)—along with provision or availability of online resources explaining how well these treatments might work, their side effects, and their bearing on other life concerns (issues)—lead to greater alignment and a stronger sense of self-efficacy?

Minimally disruptive medicine

Living with even one chronic health condition like RA places a significant burden on someone. It entails lifestyle changes, enlistment of support, psychosocial stress, dealing with symptoms, treatment management, potential changes in health status, among other consequences. One's life changes. This burden is only increased if a person has to deal with multiple chronic conditions, a situation faced by 1 in 4 American adults in 2010 (Ward, Schiller, & Goodman, 2014). Clinical practice guidelines and the best available evidence may suggest an optimal care plan, but implementation of that plan may exceed the capacity of someone to execute it in the realty of their day-to-day life. Such is the rationale behind minimally disruptive medicine (MDM) (Leppin et al., 2015). It promotes a process by which the "right care" can be identified and implemented, by working "with patients and caregivers to design care that advances patient goals with the smallest possible healthcare footprint on their lives." The theory goes that the best care is that which the patient can truly integrate into their life. It requires that the capacity of the patient to integrate that care be expressly acknowledged (Leppin et al., 2015).

In principle, resources that help people understand the potential impact of treatment choices on their lives can help facilitate MDM. If a patient understands the daily routine changes or potential side effects that a therapy might entail in balance with its putative benefits, they might be able to more confidently accept or reject it. MDM posits as an explicit "care model" with specific inputs and dimensions along which patient needs and care options can be analyzed (Leppin et al., 2015), including the patient's capacity to adopt care and the burden that specific care options necessitate. These could be (in a limited way) directly translated into a user interface. Unfortunately the prototype does not incorporate MDM principles in this way. Nevertheless, it can be understood in the context of MDM: For a patient, exploring evidence

about a treatment option that *reveals* information about that option's burden, can participate in finding what MDM calls the "right care."

Of course, the "right care" may not be the most effective therapy for a particularly disease management outcome. Considering RA, for example, it may be that the "ideal" goal of treatment with a DMARD is relatively aggressive treatment to a target of disease remission. (That would, of course, be alongside a wide range of other accommodations that a patient needs to make in order to live with RA.) Perhaps the "best" medication makes this particular patient sick, and interferes with their ability to work. If they cannot work, they cannot afford to live and eat. The "right care" for that person might not be the best medication to treat-to-target, but an alternative that achieves the best possible disease mitigation and permits the patient to keep working.

Patient reported outcomes

Quoting guidance from the U.S. Food and Drug Administration, Deshpande et al. (2011) say that a PRO is "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else." PROs represent an opportunity to deepen patient-centeredness, by acknowledging a patient's experience as *they* experience it—though mediated by an instrument of some sort, such as a visual pain scale or a categorical questionnaire response.

Measuring outcomes in a disease as complex and encompassing as RA is challenging. There is no single measure. *Success* is a combination of effects: Patient self-report of *feeling well* (e.g. less pain, less stiffness, and less fatigue), adequate function (e.g. capacity to perform tasks of everyday life, work, and leisure), physician assessment (e.g. their patient's overall health and of tender and swollen joints), lower disease activity (e.g. low levels of inflammatory factors in blood assays and evidence of slowed disease progression by radiography), and tolerance of therapy (e.g. effective lifestyle changes, no evidence of toxicity or unacceptable side effects from medications)—and more. Many of the existing measures—discussed in Chapter 2—may not be intuitively understandable; for instance, what does it *feel* like to take oral methotrexate and have a 20% improvement in a combination of tender and swollen joint counts, self-reported pain, and physician assessment of disease activity? That is an example of a typical outcome measure in research on RA treatments—a composite measure called *ACR20 (Boers et al., 2014)*—and sufficient for a pharmaceutical company to claim that a drug effects a "reduction in the signs and symptoms of RA" (Food and Drug Administration, 1999). Not only may such measures be difficult for non-specialists to understand, or difficult to translate into a meaningful form for patients, but they may not accurately reflect their priorities either (Rendas-Baum et al., 2014). OMERACT, an organization dedicated to systematization of methodologies and measurement of outcomes in RA research, has made development of instruments to soundly collect PROs a priority, and says that "capturing the patient perspective is an important part of research since the objective is to ultimately improve clinical outcomes for patients" (Boers et al., 2014). However, just as with other outcome measures, PROs are subject to the same rigor necessary to validate them—a non-trivial exercise. A PRO must be proven to consistently (from patient to patient)

measure an underlying concept (e.g. fatigue) reliably, be sufficiently sensitive to change to matter to patients, and reasonably easy to measure.

PROs bring with them the promise of assessing and communicating evidence about RA medications in terms that ordinary people might understand. They may focus research on what patients find valuable, and help answer questions that really matter to people—for example, whether interventions have a meaningful effect on sleep quality (J. R. Kirwan et al., 2009).

Summarizing a systematic review of PROs in RA by Kalyoncu et al. (2009), Gossec et al. note that only 4% of 50 trials included PRO data on fatigue, 2% on sleep, 2% on work and social life (and even then, only on productivity loss), and none on overall well-being, even though these are important to patients (Gossec et al., 2015). Absent such direct measures, additional effort is necessary to translate the "meaning" of existing measures which are perhaps suited to researchers but not so well to patients. As this project initially set out to communicate evidence about the effect of RA medications on patient experiences of domains like exercise, sleep, work and overall well-being, the lack of both PROs *or* research on these topics in general was an acute practical problem.

Decision aids and similar instruments

Decision aids (DAs) and decision support instruments (DESIs) are artifacts designed to help people understand choices and make decisions. They may take forms as diverse as brochures, checklists, tables, narrative videos, or interactive digital applications. Patient-facing medical decision aids are often static, paper artifacts. They typically are text-based, with some graphics or illustrations, and occasionally data visualizations. DAs may be used to explain any kind of intervention or health choice, such as a screening test or a treatment option, or several of these, usually oriented to a specific decision that a patient might face—for example, deciding whether to take a screening test, or starting a new medication. They may precisely describe a decision-making process, or they may simply furnish information *useful* in a decision-making process. They might include a facility for explicitly capturing patient values and preferences in the context of presented options (Stacey et al., 2012). Some DAs are designed for use during clinical encounters, sometimes to stimulate conversations about intervention options (Barton et al., 2014; Montori, Breslin, Maleska, & Weymiller, 2007). Others are designed for evaluation at home, outside the clinic.

Evidence underlying DAs may come from sources as diverse as clinical practice guidelines, clinical trials, epidemiological studies, and patient stories or interviews. Selection of evidence for DAs is inherently editorial; the International Patient Decision Aid Standards collaboration (Glyn Elwyn et al., 2006) maintains quality measures for decision aids which state that DAs ought to "base their information on comprehensive, critically appraised, and up-to-date syntheses of the scientific evidence…this refers to the information about the various relevant options, and about the descriptions and likelihoods of those options' effects on the outcomes of most importance to patients" (Montori, LeBlanc, Buchholz, Stilwell, & Tsapas, 2013). Patient-reported

outcomes (PROs) might enhance DAs that otherwise would be "limited" to difficult-to-translate findings from research.

Like any artifact that patients might encounter that confers information about medical choices, DAs are subject to bias due to presentation choices. IPDAS guidelines say that DAs should be balanced, and there is evidence that certain ways of presenting information are more likely to be perceived as balanced (Abhyankar et al., 2013). However, most presentation modes are still subject to interpretive bias on the part of viewers; for instance, there are many ways in which quantitative information about risks can be misinterpreted (Trevena et al., 2013).

Especially with paper-based, non-interactive DAs, there are limits to what can be presented. With limited space, editorial decisions about what to include and how to include it are particularly important. This may be a strength—research can reveal that which is most valuable to include—but also a weakness: New treatment options, new PROs that could be even more valuable than earlier evidence, or new evidence in general, cannot be easily added. If a team had been convened to review evidence and build a DA, but then disbanded after the DA was printed, it may have a limited shelf-life. Unfortunately, little research has been conducted on DAs built for use online (Hoffman et al., 2013), and especially little on resources *like* DAs that one might encounter in typical Google searches. DAs or similar resources are generally intended and especially valuable for decisions in which there is likely to be a state of clinical equipoise, when decisions are very sensitive to patient preferences, when potential risks are serious, or when options are unfamiliar.

The focus of this thesis project is non-clinical resources (i.e. for use outside of clinical encounters, at home, perhaps without clinician involvement at all). However, there may be special value in resources that are used in clinical encounters—should clinicians feel prepared to use them (Abadie et al., 2009), and time permits. As observed by Elwyn et al. (2013), When the paper DA is handed to a patient, it may "signal the clinician's respect for patients as relevant contributors to the process of making decisions" and a shift in power from doctor to patient. Perhaps if a digital DA were running on a tablet computer and handed to a patient, it could function similarly. It is unlikely that the same digital resource discovered and accessed at home could function similarly, though it vests similar agency for exploration and navigation of medical evidence in the patient. In either case, DAs may enhance decision making by empowering patients to be more knowledgeable participants in decisions, or as artifacts to support better conversations. They have been shown to have positive benefits in terms of knowledge and comfort with medical decisions, but not necessarily better intervention adherence or health outcomes (Stacey et al., 2012). Ultimately, their efficacy has not only to do with their design, but with *how* they are used in the social processes of patient-provider communication and decision making.

Option grids

Not all questions people have when they face a complex medical decision are general in nature, as broad as "how well does this medication work?" They may be very specific and intimately tied to the choice at hand. Intuitively, if one

Figure 2-20
An option grid.

option grid

Breast cancer: surgical options

Use this **Option Grid™** decision aid to help you and your healthcare professional talk about how best to treat breast cancer.

Frequently Asked Questions ↓	**Lumpectomy with radiotherapy**	**Mastectomy**
What is removed?	The cancer lump is removed, with some surrounding tissue.	The whole breast is removed.
Which surgery is best for long-term survival?	Survival rates are the same for both options.	Survival rates are the same for both options.
What are the chances of cancer coming back in the breast?	Breast cancer will come back in the breast in about 10 in every 100 women (10%) in the 10 years after a lumpectomy. Recent improvements in treatment may have reduced this risk.	Breast cancer will come back in the area of the scar in about 5 in every 100 women (5%) in the 10 years after a mastectomy. Recent improvements in treatment may have reduced this risk.
Will I need more than one operation on the breast?	Possibly, if there are still cancer cells in the breast after the lumpectomy. This can occur in up to 20 in every 100 (20%) women.	No, unless you choose breast reconstruction
How long will it take to recover?	Most women are home within 24 hours of surgery.	Most women are home within 48 hours after surgery.
Will I need radiotherapy?	Yes, for up to six weeks after surgery	Radiotherapy is not usually given after mastectomy.
Will I need to have my lymph glands removed?	Some or all of the lymph glands in the armpit are usually removed.	Some or all of the lymph glands in the armpit are usually removed.
Will I need chemotherapy?	You may be offered chemotherapy, but this does not depend on the operation you choose.	You may be offered chemotherapy, but this does not depend on the operation you choose.
Will I lose my hair?	Hair loss is common after chemotherapy.	Hair loss is common after chemotherapy.

Editors: Glyn Elwyn, Lisa Caldon, Kari Rosenkranz, Dale Collins Vidal, Marie-Anne Durand, Stephanie Sivell, Malcolm Reed
Evidence document: http://optiongrid.org/admin/resources/grid/evidences/9.pdf?x=WBZ1nd4YE
Publication date: 2015-05-27 **Expiry date:** 2016-03-14 **ISBN:** 978-0-9550975-6-0 **License:** CC BY-NC-ND 4.0 (International)
Download the most current version from: http://optiongrid.org/option-grids/grid-landing/9

of the options is major surgery on one's knee, a natural question is how the surgery will affect walking, and how long recovery may take. An option grid (OG) is a special kind of DA (see Figure 2-20) that answers specific common questions people have, for each of several courses of action in a medical decision (Elwyn et al., 2013). Like many DAs, option grids are designed to be used during the clinical encounter, to help answer questions that arise in conversation between provider and patient. An advantage of OGs is that they contextualize evidence and are tailored to pointed questions that real patients ask. They summarize the best available evidence to directly and clearly answer those questions. In that way, they differ from the prototype, which is intended to accommodate general evidence, and present it in a general way. Option grids are well-designed to illuminate such direct answers. As Elwyn et al. (2013) note, OGs make options concrete and literally visible (textually), illuminating

information that supports decisions by "[focusing] on the attributes that support relevant comparisons." They have been created for dozens of medical decisions, from amniocentesis tests to insulin treatment options for type I diabetes in children (The Option Grid Collaborative, n.d.). In some cases, they contain a breadth and depth of information that would be nearly impossible to spontaneously and completely cover in a clinical encounter, without the aid as a prompt.

Like most paper DAs, OGs have a few weaknesses. One is that they are only current when they are published, up to date with whatever evidence was used to create them. They are static and cannot accommodate options—such as new treatments or diagnostic choices—that were not explicitly included at the time they were made. Similarly, they only answer a limited set of questions deemed important by the authors—as well-founded and valuable as they may be to the practical decision that the patient faces.

Option Grids are built around evidence documents that show which sources were used to answer the questions. They are also distributed according to a Creative Commons attribution license which means that they can be widely accessed (The Option Grid Collaborative, n.d.). Both of these features mean that they function as a way to improve transparency and access to evidence that can support complex medical decision making.

Mayo Clinic decision aids

Several DAs developed by the Mayo Clinic Shared Decision Making National Resource Center—including one for statins (cholesterol-lowering drugs), type 2 diabetes medications, osteoporosis medications, and antidepressants—influenced the design of the prototype. They are typical of DAs in that they cannot arbitrarily accept new evidence. The aids are accompanied by a bold disclaimer: "Caution: This application is for use exclusively during the clinical encounter with your clinician," clearly indicating that they are intended for use during the clinical encounter. Both the antidepressant and diabetes DAs are static: They feature a fixed set of options and fixed design (see Figures 2-21 and 2-22). Although a digital version of the diabetes DA is available, it simply replicates the paper cards. Users cannot interactively query the DA to limit to a medication suitable given a certain comorbidity, for example. The Bone Health Choice (osteoporosis) and Statin Choice DAs are digital, data-driven resources. They are tailored to the patient's specific situation to estimate the baseline risk of a negative health outcome—a heart attack in the case of Statin Choice. The putative benefit of the intervention—prevention of a heart attack—is expressed in the context of that baseline risk (see Figure 2-23). However, they also do not accept new evidence. In both, the intervention is a class of medications (e.g. the only "choice" is taking a statin or not)—reasonable since the medications may be equally effective for the particular outcome of interest—rather than individual medications. In that way, the DA cannot be tailored, to limit to a medication suitable given a certain preference or comorbidity.

These DAs feature an *issue*-centric design. Data are presented in the context of issues relevant to people facing the particular decision, or that are unique to the treatment options, alongside general concerns like side effects, cost, and daily routine, as the Diabetes Medication Choice image in Figure 2-22

illustrates (Mayo Clinic Shared Decision Making National Resource Center, n.d.-b). Like the frequently asked questions in OGs, these issues are germane to the choice being made. For example, a single potential serious risk associated with the choices may be elevated to the status of an issue and become its own card–such as the risk of low blood sugar with antidiabetic medications. Similarly, top antidepressant side effect concerns–such as sexual side effects and weight gain–are elevated to the status of a "first-class" issue in the Depression Medication Choice DA. In the paper, card-based DAs, medications are generally found in the same vertical position on each card, so that multiple issue cards can be used as "lenses" to compare medications side by side. Or, a single card can be used to compare multiple options along a single dimension. Visually, these DAs employ symbols, graphical representations, relatively little text, and visualizations like icon arrays that may ensure more effective risk communication and comprehension for lower literacy patients (Galesic et al., 2009). With a clinician present, patients can ask for clarification or discuss anything they do not understand. The DA is an artifact that facilitates conversation about the choices and issues (Montori et al., 2007).

As these DAs are designed for use during clinical encounters, practitioners must learn about them, trust them, and find a way to integrate them into usual care. And, as Breslin et al. note in their discussion of the diabetes DA, "a conversation-based decision aid relies heavily on the communication skills and knowledge of the clinician, attributes that may vary greatly across clinicians" (Breslin, Mullan, & Montori, 2008). Omitting information that would be necessary to make a fully informed decision is a clever and deliberate design choice, given the conversation philosophy–the clinician and patient together *complete* the DA. The authors note that a "stand alone decision aid" (such as the thesis prototype) could "overwhelm patients with complex information

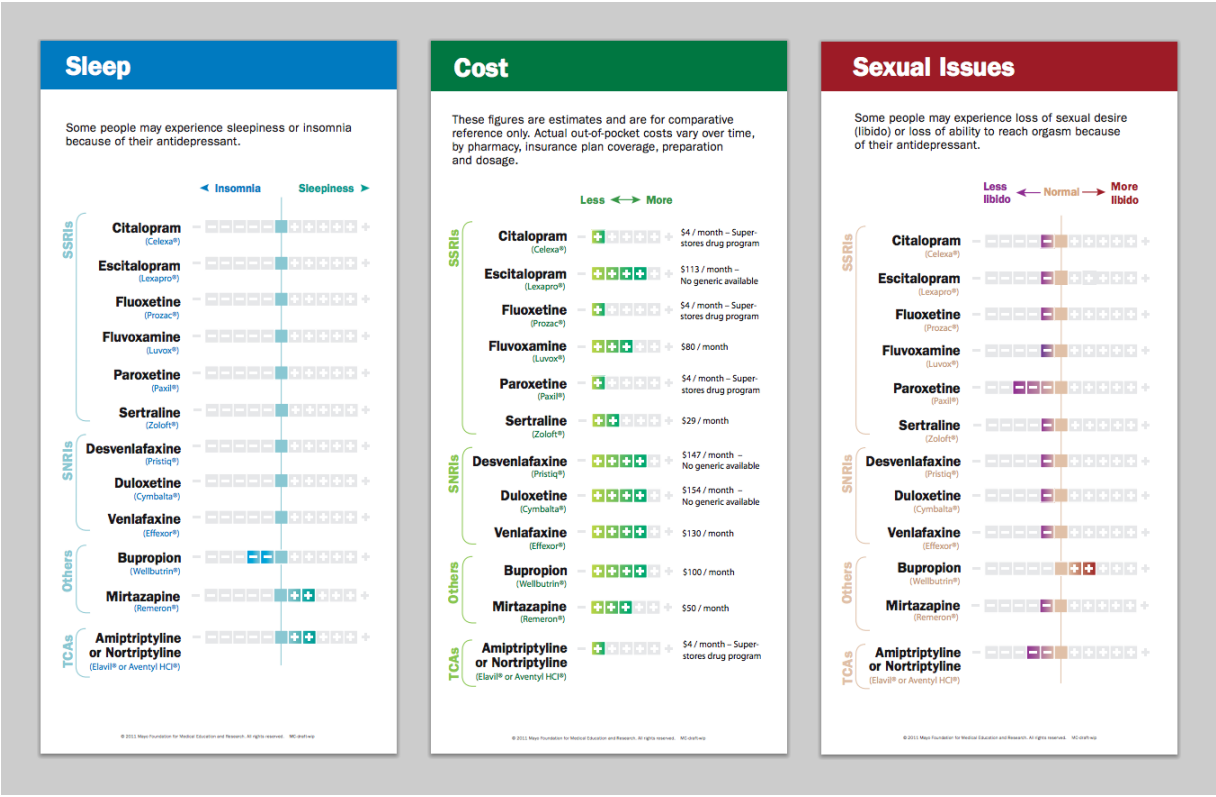


Figure 2-21
Mayo Clinic antidepressant decision aid. These are a subset of the "issue" cards that comprise the decision aid.

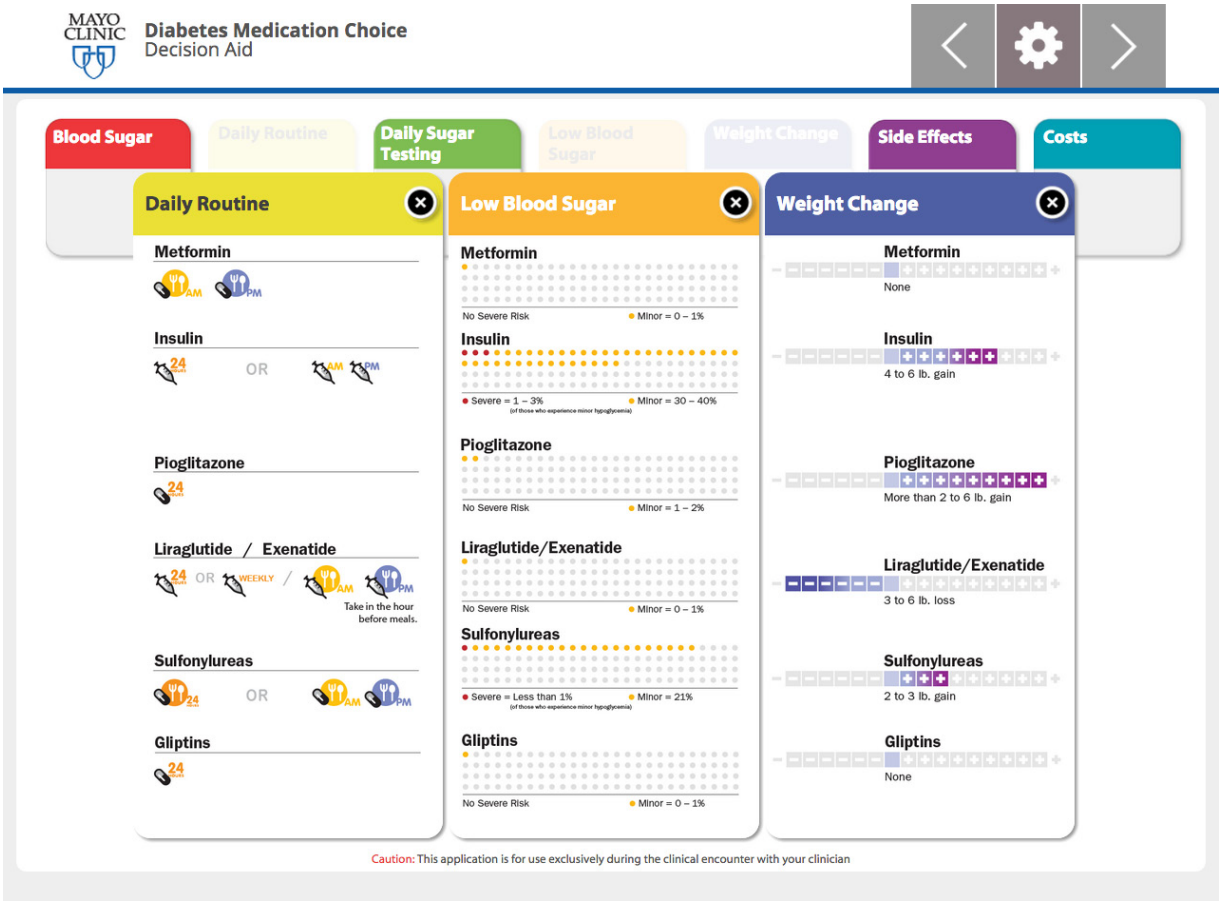


Figure 2-22
Mayo Clinic diabetes medication decision aid. This online version replicates the equivalent printed cards.

and deprive patients of the clinician's stories of other patients' experiences and their expert judgment," which is a legitimate possibility.

Cochrane methotrexate decision aid

http://musculoskeletal.cochrane.org/decision-aids/

One of several DAs published by the Cochrane Collaboration, this methotrexate DA (see figure 2-24) is a paper-based, linear, narrative document explicitly designed for the choice explicated in its title: "Should I take methotrexate (Rheumatrex) alone or with other disease-modifying anti-rheumatic drugs for rheumatoid arthritis?" (Rader et al., 2011). It is intended for people who have been taking methotrexate but whose RA has not improved. It features explicit values elicitation (i.e. an exercise to help people identify "what matters most" to them), icon arrays to communicate risk information (about benefits and side effect risk), a knowledge quiz, and specific "next step" options (including "I will take methotrexate with other disease-modifying anti-rheumatic drugs"). This document is bespoke and up to date only with the evidence cited (a single source), and offers no specific side effect risk information or information about the "other" DMARDs. In this way, it feels incomplete even though it points patients towards a very specific decision to which the missing information is germane. The evidence on this DA suggests that few people who have not improved already with methotrexate will improve after another 1 to 2 years,

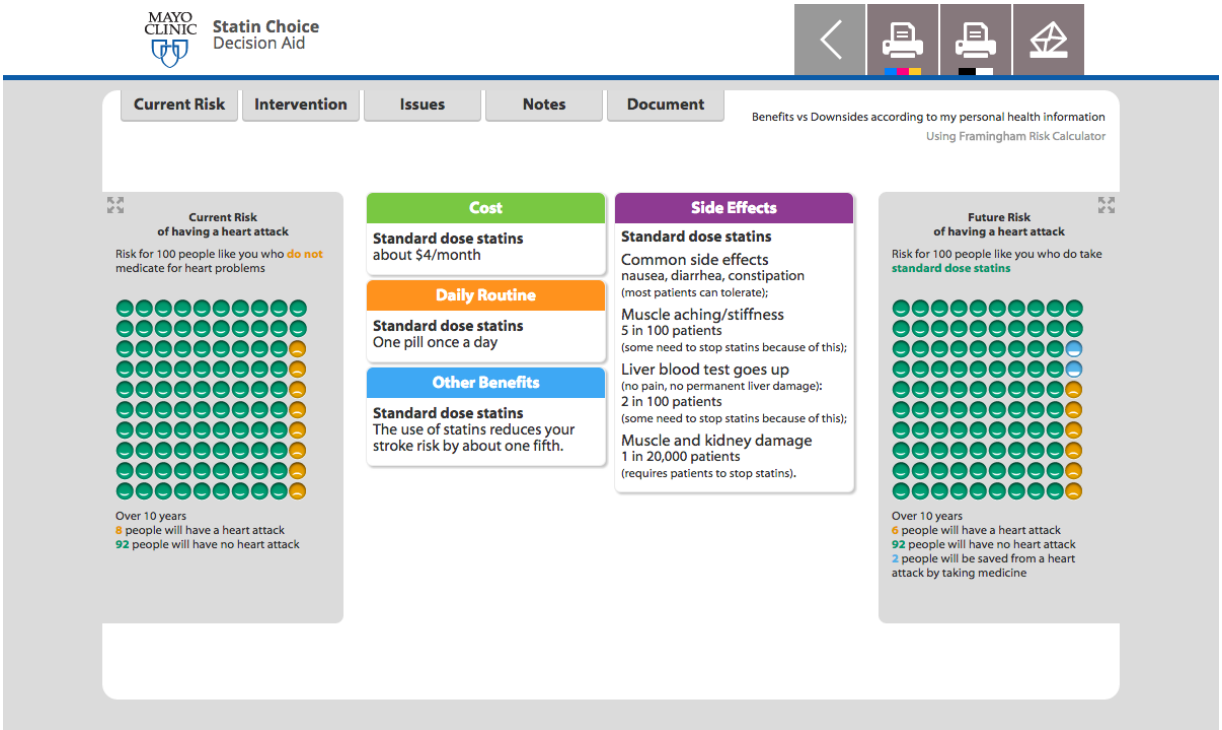
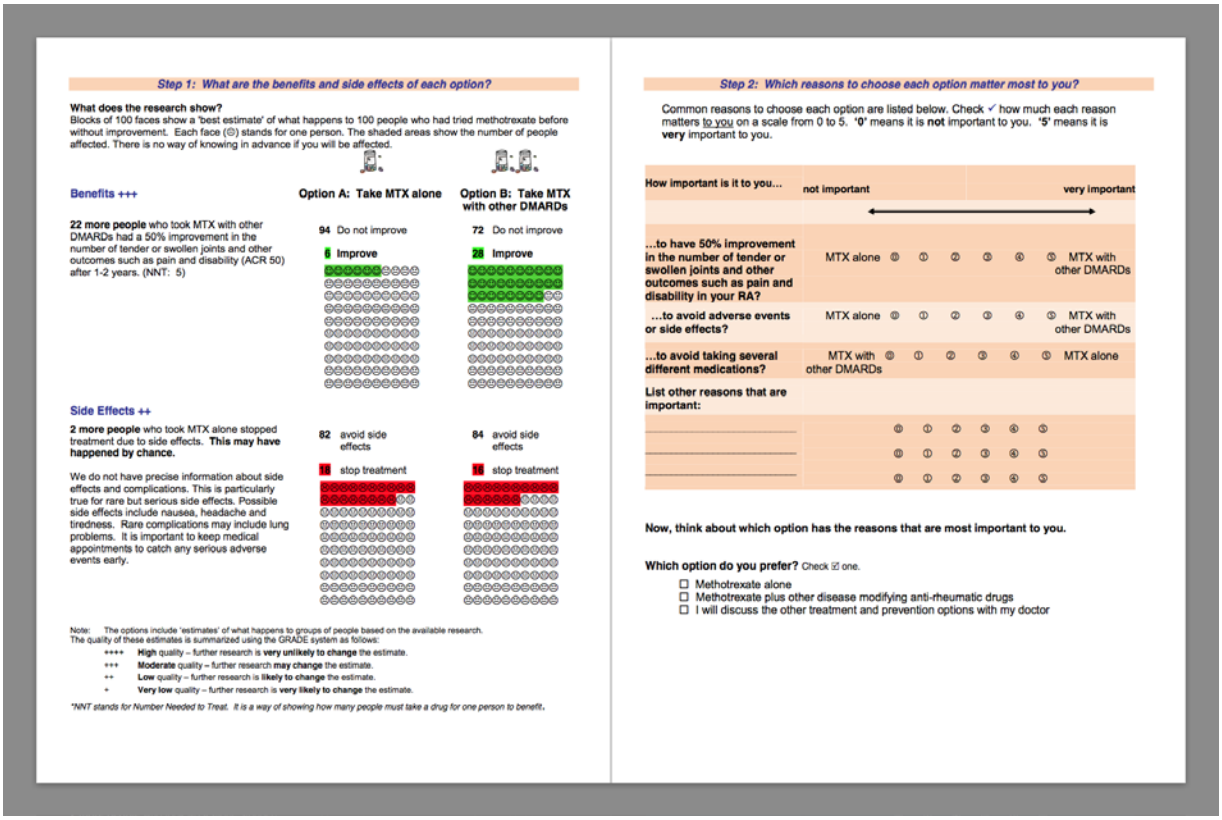


Figure 2-23
Mayo Clinic statin decision aid. In this interactive decision aid, issue cards are supplemented by tailored risk visualizations (icon arrays) showing the relative risk of having a heart attack without taking a statin versus taking a statin.

while people who take methotrexate with another DMARD are more likely to improve, while there is no difference in terms of the estimated number of people who would stop treatment due to side effects with either therapeutic option. However, parts of this DA may be confusing. For instance, in the values elicitation exercise, it suggests that people who consider avoiding adverse events or side effects *unimportant* would likely prefer to use methotrexate alone, while people who consider it *very important* to avoid adverse events or side effects might prefer methotrexate with another DMARD—even though the DA itself says that people on the combination therapy do not stop treatment at a rate any different than those who take methotrexate alone, and even though a new medication introduces the risk of new side effects that might be unfamiliar. This suggestion may be because RA disease progression itself may result in serious adverse events, although that is not clear from this exercise.

Like most DAs, because this one is paper-based, it cannot be kept up to date with new evidence. It also cannot itself be used to explore the other DMARD options. While the Cochrane group publishes similar DAs for a few other DMARDs, each new one needs to be crafted by hand.



Step 1: What are the benefits and side effects of each option?

What does the research show?
Blocks of 100 faces show a 'best estimate' of what happens to 100 people who had tried methotrexate before without improvement. Each face (☺) stands for one person. The shaded areas show the number of people affected. There is no way of knowing in advance if you will be affected.

Benefits +++

Option A: Take MTX alone

94 Do not improve

6 Improve

Option B: Take MTX with other DMARDs

72 Do not improve

28 Improve

22 more people who took MTX with other DMARDs had a 50% improvement in the number of tender or swollen joints and other outcomes such as pain and disability (ACR 50) after 1-2 years. (NNT: 5)

Side Effects ++

2 more people who took MTX alone stopped treatment due to side effects. **This may have happened by chance.**

We do not have precise information about side effects and complications. This is particularly true for rare but serious side effects. Possible side effects include nausea, headache and tiredness. Rare complications may include lung problems. It is important to keep medical appointments to catch any serious adverse events early.

82 avoid side effects

18 stop treatment

84 avoid side effects

16 stop treatment

Note: The options include 'estimates' of what happens to groups of people based on the available research. The quality of these estimates is summarized using the GRADE system as follows:

++++ **High** quality – further research is **very unlikely to change** the estimate.
+++ **Moderate** quality – further research **may change** the estimate.
++ **Low** quality – further research is **likely to change** the estimate.
+ **Very low** quality – further research is **very likely to change** the estimate.

**NNT stands for Number Needed to Treat. It is a way of showing how many people must take a drug for one person to benefit.*

Step 2: Which reasons to choose each option matter most to you?

Common reasons to choose each option are listed below. Check ✓ how much each reason matters to you on a scale from 0 to 5. **'0'** means it is **not** important to you. **'5'** means it is **very** important to you.

How important is it to you...	not important					very important	
...to have 50% improvement in the number of tender or swollen joints and other outcomes such as pain and disability in your RA?	MTX alone	0	1	2	3	4	5 MTX with other DMARDs
...to avoid adverse events or side effects?	MTX alone	0	1	2	3	4	5 MTX with other DMARDs
...to avoid taking several different medications?	MTX with other DMARDs	0	1	2	3	4	5 MTX alone
List other reasons that are important:							
_____		0	1	2	3	4	5
_____		0	1	2	3	4	5
_____		0	1	2	3	4	5

Now, think about which option has the reasons that are most important to you.

Which option do you prefer? Check ☑ one.

- ☐ Methotrexate alone
- ☐ Methotrexate plus other disease modifying anti-rheumatic drugs
- ☐ I will discuss the other treatment and prevention options with my doctor

Figure 2-24
Cochrane methotrexate decision aid. These are the central 2 of 4 pages of this 4-step linear DA, which also includes an introduction, explicit knowledge quiz, and next steps section.

Clinical practice guidelines

A clinical practice guideline (CPG) is in principle an evidence-based document that makes recommendations methods for diagnosis and treatment of patients with a certain illness (Krahn M & Naglie G, 2008), although not uniformly rigorous in their formulation or recommendations (Guyatt & Vandvik, 2013). A CPG principally designed for use by clinicians. Recommendations may cover domains like diagnostics, treatment strategies, dosing, follow-up, patient engagement techniques, and more. A CPG documents consensus among specialists based on a review of evidence (Weijden et al., 2010), and indeed may be vehicles for the *standardization* of evidence-based care: If they are thorough, up to date, and clinicians do their best to follow guidelines, they may help the best care to be delivered to all patients. However, they may not explicitly include PROs or include much discussion of patient beliefs, needs, and preferences and the central role they may play in effective care (Krahn M & Naglie G, 2008). In such cases, CPGs may appear to be more focused on treating the disease rather than the patient, though clinicians can apply recommendations in the context of their patient‘s individual needs.

With respect to design, CPGs are often long, narrative documents with some diagrams and tables, but not machine-readable data nor digital and interactive. This can present a problem for practitioners who may need to rapidly look up a recommendation germane to a patient, to help them make an evidence-based decision (Vandvik et al., 2013). The Institute for Clinical Systems Improvement (ICSI) guideline for treatment of adult depression in primary care

(i.e. intended for primary care physicians) is 88 pages long—plus a bibliography (ICSI, 2013). The more compact American College of Rheumatology 2015 guidelines for the treatment of RA are 20 pages, plus references (Singh et al., 2016). Because CPGs have actionable information that impacts patient care, it is reasonable to believe that they could be useful sources of evidence for patient-facing resources (Raats, van Veenendaal, Versluijs, & Burgers, 2008). That the actionable recommendations and data in guidelines may be "locked up" in lengthy text and not machine-readable parallels the situation with much evidence, and in consumer health resources. Guyatt and Vandvik describe key challenges in CPG creation and use in a 2013 article called "Creating clinical practice guidelines: Problems and solutions," saying:

> Rigorously developed guidelines face other problems of optimal usefulness. Newly published evidence may render recommendations outdated shortly after publication. Clinicians cannot easily access many recommendations published in lengthy PDF formats and unavailable in electronic medical records at the point of care or in devices such as smartphones and tablets. Finally, most guidelines cannot be directly used for shared decision-making.
>
> Solutions to this latter set of challenges (ie, adaptation, updating, access at point of care, use in shared decision-making) may be imminent. Electronic authoring tools can facilitate both a rigorous, structured approach to guideline development and local adaptation and updating. The output of such an electronic authoring process can appear on smartphones and tablets and be easily integrated into electronic medical records.
>
> (Guyatt & Vandvik, 2013)

These arguments very closely mirror the rationale for the prototype. Fortunately, efforts are underway to demonstrate solutions to these problems.

MAGIC & SHARE-IT

There are examples of pragmatic data-driven evidence-based electronic tools—including decision aids—that begin to address the problems noted by Guyatt and Vandvik. One constellation of projects investigating creation and implementation of evidence-based clinical practice guidelines in the context of shared decision making stands out. The MAGIC (Making GRADE the Irresistible Choice) collaboration "was established in 2010 to facilitate the authoring, dissemination and updating of trustworthy CPGs" and, following a rigorous design process, has demonstrated working applications (Kristiansen et al., 2014; Vandvik et al., 2013). The data-driven MAGICapp system ingests summaries of evidence for CPGs (including data, not just recommendation text) and makes them available in electronic version of CPGs; these electronic CPGs favor a certain presentation of the evidence and recommendations but are not limited to that presentation. The guidelines can be accessed on a desktop computer, a mobile app, downloaded as a PDF (not dissimilar from a traditional CPG), and they are even made available for export in JSON, readily used by the prototype or in other online resources. These are indeed the benefits of starting with machine-readable data. Furthermore, the collaboration has even developed

tools under the banner "SHARE-IT" that can facilitate "semi-automated production of a large number of decision aids" that use the very same evidence base as the CPG (Agoritsas et al., 2015). These decision aids are interactive; they support exploration of one finding at a time, with progressive disclosure–for example, one can look at the absolute risk (chance of benefit) of an intervention

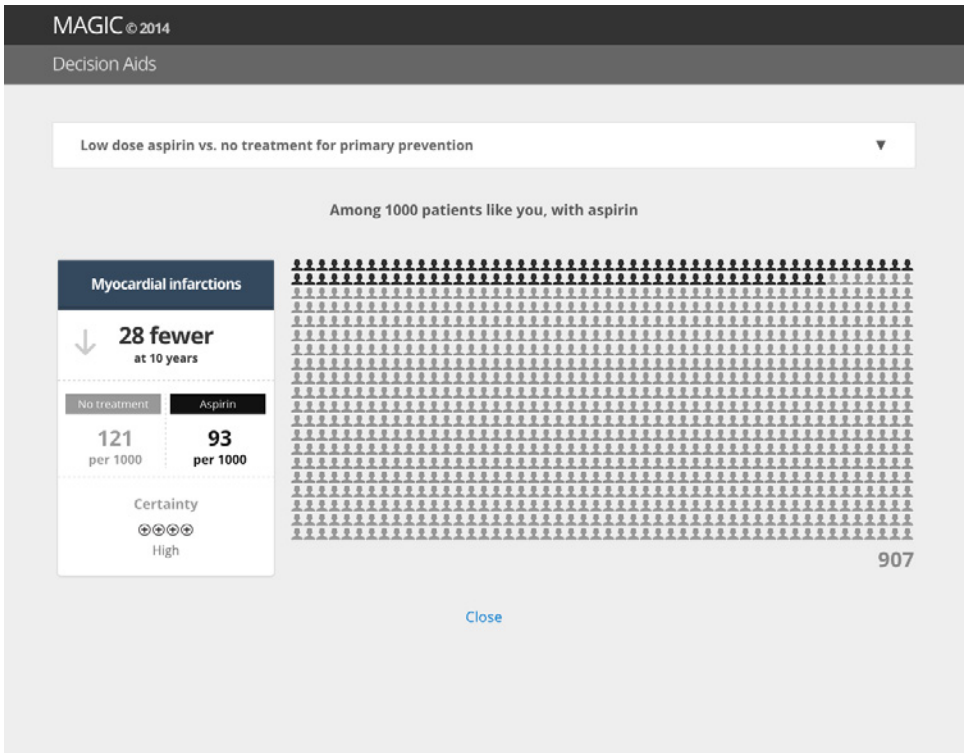


Figure 2-25
A MAGIC decision aid. This interactive icon array visualization facilitates comparison of the relative risk of having a heart attack taking low dose aspirin versus not taking it.

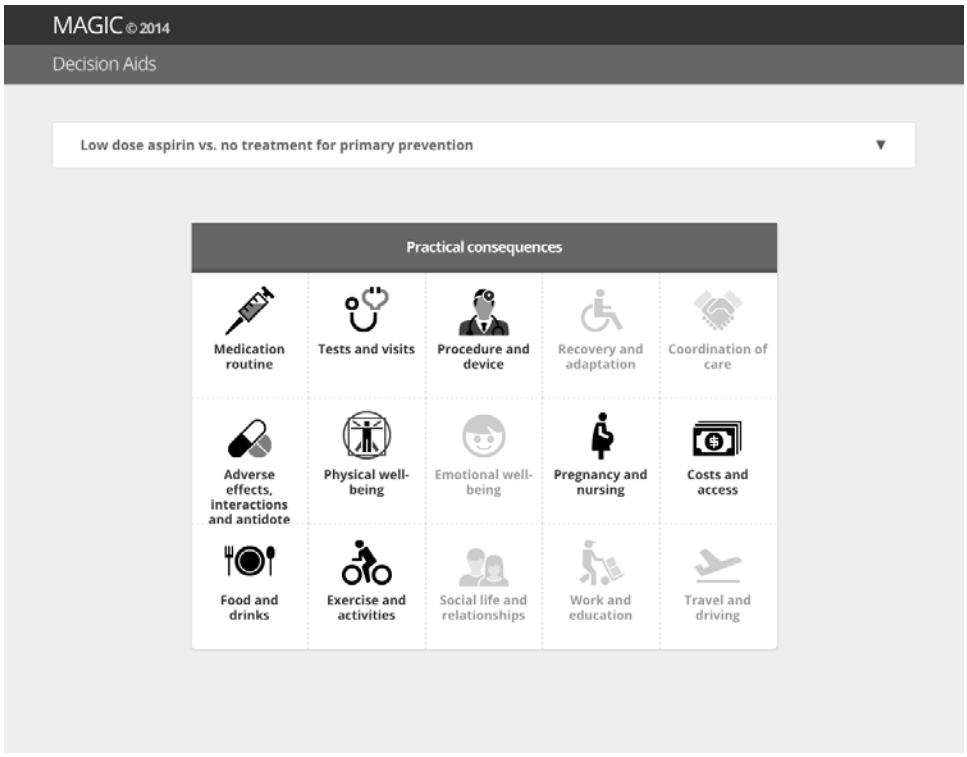


Figure 2-26
"Practical" issues in a MAGIC decision aid. These are similar to the issue cards in Mayo Clinic decision aids, and the "basic issues" feature of the *Navigator* prototype.

as a figure alone or including an icon array visualization (see Figure 2-25). These DAs feature a "practical" issues screen (see Figure 2-26) that can provide context around questions about the life impact of choosing one intervention or another, not dissimilar to the Mayo Clinic decision aids and the RA Choice decision aid. More than a dozen electronic CPGs are publicly available, some with accompanying decision aids (see the list of currently available CPGs at magicapp.org, MAGIC, n.d.).

These semi-automatically generated DAs are—in fact, the entire system that produces them—philosophically aligned with the thesis project in that they demonstrate how machine-readable data and systematized tooling can be used to produce *multiple* updatable resources for different audiences, rather than paper-based bespoke resources that quickly become dated. One can get the data into the system, and relatively easily build resources for different audiences that use those data. Like the prototype built for this thesis project, the MAGIC DAs and CPGs could share the same evidence, be continuously updated, and complement one another outside of and in the clinical encounter.

Chapter 3

Prototype

A platform for public engagement with medical evidence

The major work of this project was development of a prototype platform on which can be built UIs for exploring medical evidence. The primary activities undertaken were:

- Researching and encoding findings from medical literature and other sources of evidence in an original data schema.
- Building a software system to process those data.
- Designing and prototyping user interfaces for exploration of those data, using the re-usable components of the software system.

The product of this work is an open-source Web-based system for encoding medical evidence about RA treatments, "translating" it to higher-level patient-centered *issues* and graphic representations, and UIs to facilitate exploration by members of the general public. Its scope is limited (it presently contains only

› A demonstration application—called the *Navigator* prototype is at **http://thesis.merges.net/navigator**

› Data are encoded in a publicly visible Google Spreadsheet at **http://goo.gl/gdbVuf**

› The prototype is open source. Its code is checked into GitHub at **http://github.com/merges/abist**

a tiny fraction of suitable evidence, and just a few incomplete user interfaces and visualizations) but it is a complete end-to-end demonstration platform. It comprises the following:

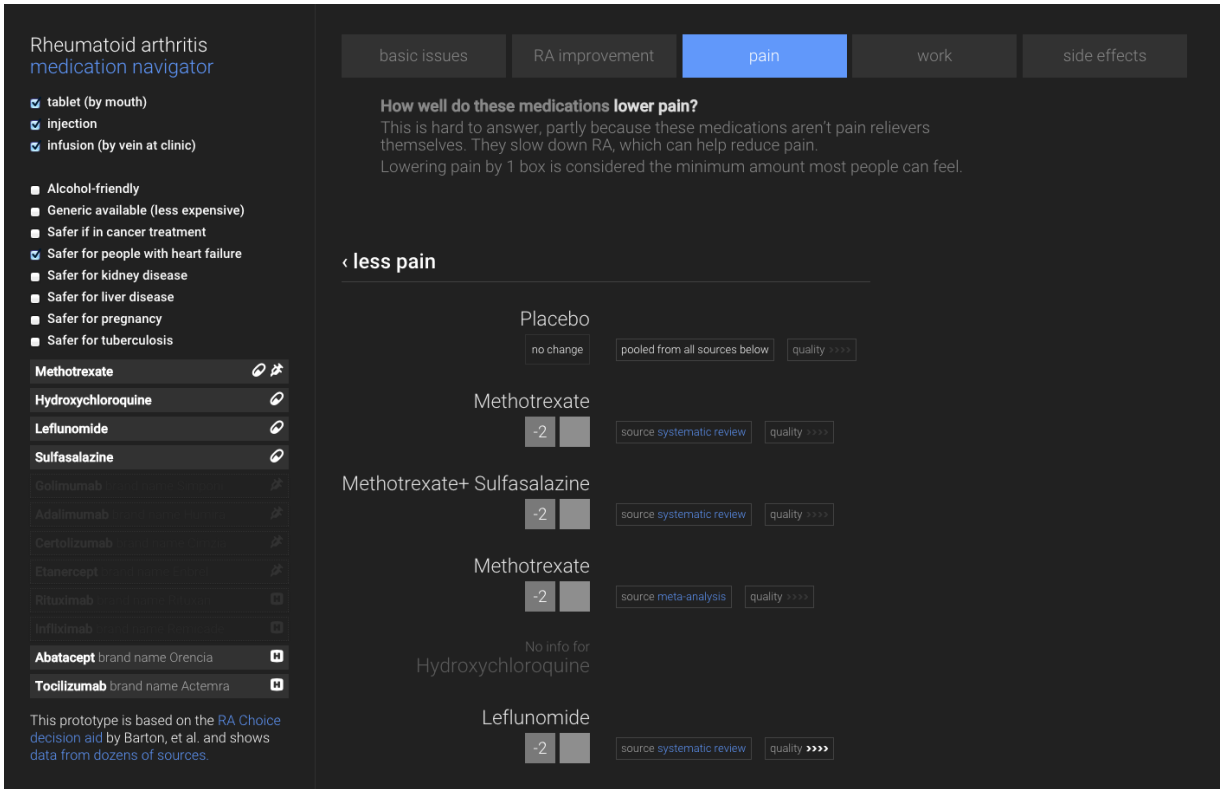
- A spreadsheet "back-end" with "findings" from various sources of evidence, including general information about 12 medications used to treat RA.
- A technical layer to interpret and transform the data, including separable modules for presentation of evidence.
- Public-facing Web applications that can be accessed from desktop computers, tablets, or mobile devices. These could be provider-facing tools to explore *all* evidence, highly narrative decision aids that incorporate evidence, tightly focused tools for looking at *one* particular kind of evidence, and so forth. Several examples were built to demonstrate how the platform works.

The *Navigator* prototype application

This is a demonstration of the kind of application that can be built using the prototype platform (see Figure 3-1). It is effectively a decision aid, though it was not developed with a deliberate methodology and would require much refinement before it would be suitable for clinical use. It instead represents a superficial example of a data-driven resource for exploring medical evidence. It supports two key activities.

Figure 3-1
Screenshot of the *Navigator* prototype application.

The first is *filtering*—choosing a medication based on individual needs or preferences (such as preferred dosage form, cost, or lifestyle considerations),



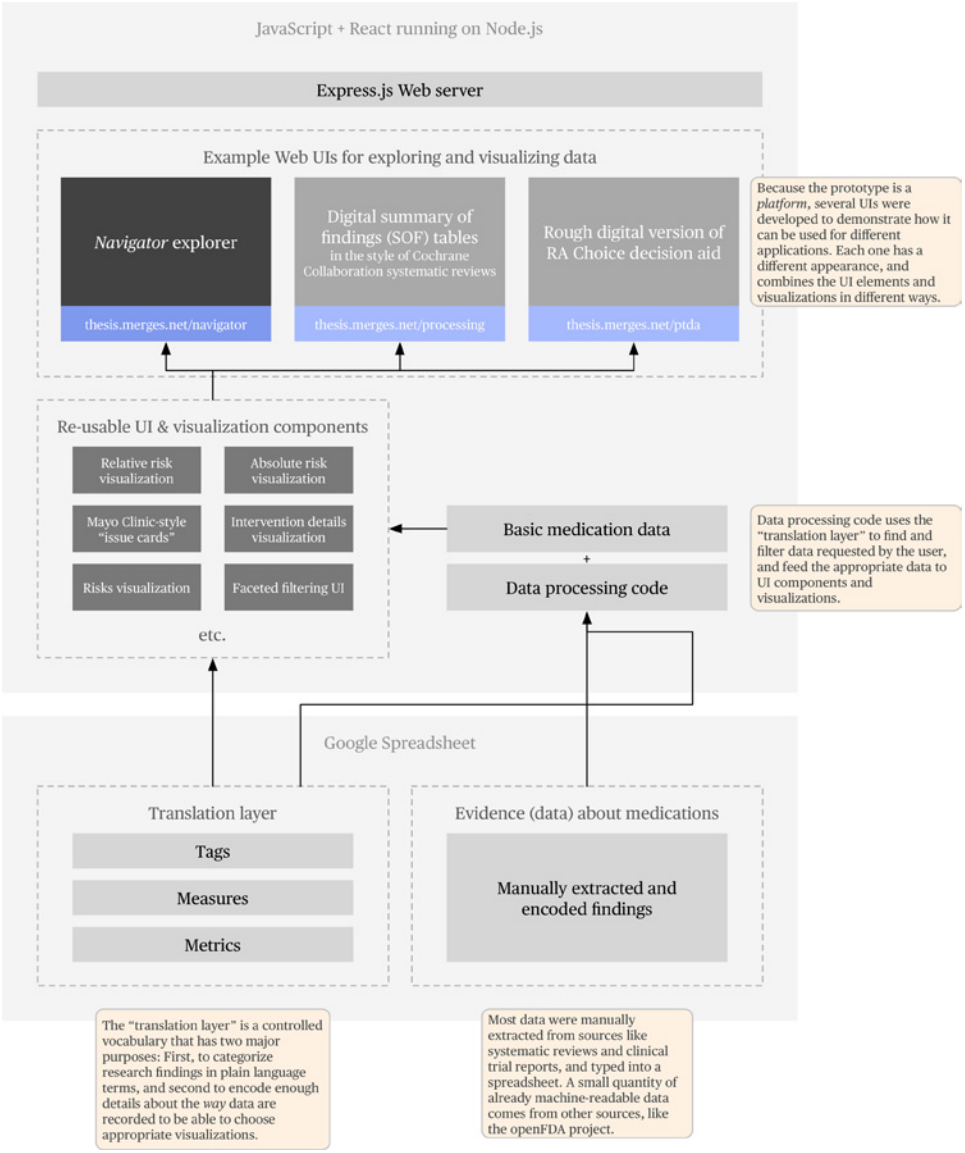


Figure 3-2
Schematic diagram of the prototype platform.

and the second is *exploration of evidence* through the lens of high-level issues (e.g. overall improvement or pain).

The whole platform, but especially the *Navigator* example, began as a simple digital version of the existing RA Choice decision aid (Barton et al., 2014). That DA's set of 12 DMARDs—methotrexate, hydroxychloroquine, sulfasalazine, leflunomide, golimumab, adalimumab, certolizumab, etanercept, rituximab, infliximab, abatacept, and tocilizumab—serves as the basis for the prototype. Some medication-specific information from the RA Choice DA was encoded and plays a role in the prototype, including typical cost, some of the common risks and warnings, and route of administration, administration routine, and delay of onset of effect—issues that were discovered by Barton et al. to be important to patients. Further information about risks was sourced from systematic reviews (Ramiro et al., 2014) (Richards, Dowell, Quinones, & Kerr, 2015). These data power the preference-driven filtering UI—so that one can limit to oral

medications, or those that are safer given a certain comorbidity, or that have a generic version available. Most of the "evidence" related to each medication is in the publicly viewable Google spreadsheet.

Figure 3-2 is a schematic illustration of the platform. It is a rough approximation of the parts of the platform and how they connect with one another. Before discussing the methods and design choices that underpin the prototype and the *Navigator* prototype, it is necessary to briefly explore a set of guiding principles that structured the development process.

Guiding principles

A suite of mutually reinforcing principles guided design, technology choice, and methods employed in developing the prototype. They were derived from a mixture of established theories, professional experience, and notions that arose from reading related literature. Since one major goal of the project was to demonstrate an extensible *platform*, an overarching principle was to ensure that a variety of products could be built with it for further evaluation—applications as diverse as a highly visual low-literacy data-driven decision aid, or a comprehensive "viewer" for all available evidence. With respect to an ideal data-driven online health resource for the general public, this author believes it would be sensitive to context, personalized, responsive to literacy and learning style, alive with the latest evidence, furnish the person making a decision with useful information, increase knowledge and confidence, and strengthen conversations between patients, caregivers, and clinicians. That tall order is beyond the scope of this project, but the prototype hopefully has taken steps towards a platform that can help build such resources.

Sufficiency

This principle conveys the idea that solutions—to technical problems, design problems, etc.—which work *well enough* should be adopted, rather than seeking a comprehensive or perfectly flexible alternative. In this way, sufficiency stands in opposition to exhaustiveness and thoroughness. It serves several conceptual and practice purposes here:

- It reflects the reality that medical decisions are made with imperfect information.
- It promotes effective work, rather than perfect work.
- It emphasizes iteration and improvement, and limits are quickly discovered.

Practicing design with sufficiency in mind resulted in a highly iterative design and development process, discussed further in a later section. Because

evaluation with real users was outside the scope of this project, it is not possible to claim that any of the prototype's "patient-facing" design choices are sufficiently effective for furnishing value to people making medical decisions. In the context of evaluation, *sufficiently effective* might mean: Do people get the *gist* of the finding, is their understanding reasonably faithful to the data, and do they report it being helpful to their decision?

- How much detail about an outcome should be revealed? Only enough to help someone compare or make a decision. Confidence intervals might only introduce confusion, for example, and if uncertainty is comparable from drug to drug, just showing a point estimate is sufficient. Indeed, in many cases uncertainty in estimates of effect has been omitted from presentation (although still encoded, and easily switched on with a few lines of code).
- How comprehensive should the data schema be? Only enough to take in evidence at the level of detail that would likely be used in a tool for the general public.
- Is a spreadsheet *good enough* as a data back-end, as opposed to a database? Yes, because it is semi-structured (i.e. data are machine-readable), and it is sufficiently rapid and easy to work with (for both the author, and for the potential audience of contributors from the medical community). It satisfies basic needs.

Common technologies and tools, which are "off-the-shelf," free to use and extend, and readily used by less-expert participants in the production and dissemination of medical evidence (researchers, physicians, writers, bioinformaticists, designers, students, epidemiologists, and so on), are more than sufficient for the creation and development of an extensible platform. These imperfect but sufficient tools were always preferred during development.

Emergence

Closely related to sufficiency is the principle of reliance on emergence, as opposed to comprehensive planning. In this context, emergence refers to design solutions that come into being as a result of iteratively working with data and developing the prototype. In effect, they are revealed as the whole is built, and cannot be defined in advance. With sufficient tools and structures—those that meet minimum requirements, but that are flexible and can accommodate change—optimal solutions can emerge from experimentation. For example, rather than adopting a specific data schema *a priori*, the current (but not final) schema "revealed itself" as development and design practiced, with new data and prototyping work continually stressing its limits. Occasionally, boundaries were discovered that meant reworking the schema, an inherent risk to relying on solutions to emerge through experimentation.

Systematization

Most design decisions were made with an eye towards systemization, in order to reflect the goal of a *platform* or *system* to support data-driven health resources for the general public. The goal was not to make a bespoke artifact that only the author could understand, but instead a shareable suite of artifacts that can be reassembled, re-used, or at least examined by interested outside

parties. Here are a few examples of how that played out during development of the prototype:

- **Technical decisions.** Contemporary, open-source technologies so that people other than the author could read, adopt, and work with them.
- **Social decisions.** Using spreadsheets, so that other people could easily be brought on board to participate in adding evidence to the prototype.
- **Development decisions.** Creating reusable components for common elements that might be re-used, like absolute risk visualizations – usable across all kinds of resources. Creation of a fairly general data schema that can be used for many kinds of outcomes, not just those in RA or those encountered during the course of research conducted for this thesis project.
- **Design decisions.** UI that can in a basic sense accommodate a continually expanding number of data points, etc. The visualizations, for example, are largely data-agnostic–for instance, they can accommodate any outcome for any medication, although they may not be optimally suited for such purposes.
- Whenever possible, a *general* and re-usable systematic approach was used instead of a bespoke process that could only be applied once to a very specific problem. For example, none of the designs implemented so far are uniquely tied to rheumatoid arthritis medications.

Elasticity

One potential shortcoming of DAs–both paper-based, and even digital examples such as the Mayo Clinic aids–is that they represent a fixed perspective on the evidence at a certain point in time. To be sure, new sweeping and conclusive data that would radically change a typical DA are not being published and verified on a weekly basis. So the *core* of a DA might remain relatively unchanged. But new data–about a specific comorbidity, for example–does come out from time to time. It would be nice if any DA that offered tailoring based on one's individual health status would immediately update after new evidence had been reviewed and approved for addition.

During the course of the prototype's development, one of the author's supervisors forwarded a new article about the safety of biologic DMARDs for patients with certain comorbid conditions (Richards et al., 2015). Relevant findings from that article were encoded in the prototype in just a matter of hours, improving the medication filtering. Similarly, new "outcome" findings (of the sort on which the prototype is based) can be incorporated by adding one or two lines to a spreadsheet. Elasticity–in this case, capacity to accommodate new evidence–was central to the prototype. There were implications in terms of design (for instance, the design could not assume a fixed number of drugs or data points or outcomes), data (the schema had to be flexible), and tooling (adding findings had to be relatively easy). The data schema that has emerged so far can accommodate *most* but not *all* outcome-type findings from medical research.

Why design for this elasticity? Consider this excerpt from the American College of Rheumatology 2015 update to their guidelines for treating RA:

> Due to rapidly evolving knowledge for the treatment of RA, some recommendations may be outdated by the time they are published due to the emergence of new evidence. Examples include new data on tapering and discontinuation of therapies in early RA and treat-to-target. The short half-life of treatment recommendations is also related to the rigorous and time-consuming process of guideline development used by the ACR, which complies with guidance from the National Academy of Medicine (formerly the Institute of Medicine) and the Council for Medical Subspecialty Societies. Additional time is also required for review and endorsement of each guideline document by ACR committees, journal reviewers and editors, and the ACR Board of Directors. However, the ACR regularly updates RA guidelines and strives to shorten the time between the end of the literature review and the publication of guidelines, to make them as relevant and current as possible.
>
> (Singh et al., 2016)

It is clear that the rigorous review process—a socially complex process involving possibly dozens of people and intensive literature review and discussion—takes time, preventing new evidence from changing guidelines as it becomes available. And it is reasonable to only want guidelines to be "rubber stamped" after a thorough review in which all of the evidence is considered together (a kind of synthesis). However, there is an advantage to all such resources being "living documents" that can grow and accommodate new evidence as it is discovered.

Reusability

Although this principle could not be fully realized, wherever possible parts of the prototype system were designed for adaptation and reuse as individual elements. Instead of designs and technical choices intimately tied to *this* prototype and RA *only,* features of the platform—the data schema, data visualization code, and so on—can be reused in other contexts. Indeed, several prototype applications have already been built with it, although the reusability of the prototype's components have not been tested by other developers.

Evidence agnosticism

In principle, it could be considered an ethical imperative that patient-facing online resources use trustworthy, high-quality or highly *certain* sources of evidence. (Some of the challenges in evaluating trustworthiness and quality are elaborated later in this chapter.) However, buttressing the principle of elasticity, the prototype is deliberately agnostic to the source and quality of evidence. It can accommodate a new finding from a potentially quite biased clinical trial, or a highly certainestimate of effect from a low-bias and well-conducted meta-analysis. It *does* support encoding of the quality of that particular finding, and links to the source material; also, reusable UI components for communicating quality and source have been built, and the example applications (including the *Navigator*) show the provenance of data. The hope is that applications built

with the platform would be able to decide on what kind of data to include. For example, it can be as simple as adding a checkbox to only show findings that come from a particular kind of source, or an express design decision to exclude certain kinds of data. In this way, the burden of *which* evidence is appropriate is shifted to humans, rather than restricted by the prototype in some technical way. However, it would ultimately be up to the designers and developers and users to decide which sources are appropriate.

With respect to communicating the certainty of evidence, the prototype so far only supports one system, GRADE (G. H. Guyatt et al., 2008). A useful future addition to the prototype would be a harmonization layer to communicate a unified notion of quality or certainty to patients, regardless of the system used to evaluate it. However, this is a difficult and unsolved problem since there are many ways in which these can be evaluated.

Harmonization, not synthesis

Synthesis—in this case, combining evidence and producing a kind of intelligent composite *whole* picture or interpretation—was an anti-principle guiding the project. While it is probably the ultimate goal of a patient-facing tool (after all, it would mean arrival at something closer to an *answer* like "this medication works best" or "this medication might work best *for you*"), it is exceedingly difficult and if poorly executed, potentially misleading and dangerous. Because the prototype is elastic, and can accommodate arbitrary new evidence, any algorithms for synthesis would be required to accommodate those data and incorporate them faithfully and accurately. There are a number of reasons why that was impractical:

- **Heterogeneity.** Evidence is heterogeneous in innumerable ways, including outcomes, measures, metrics, uncertainty around data, methods of production, populations studied, intervention details, methods of analysis, implicit or explicit biases, etc. In order to synthesize these data, explicit decisions must be made about how to address inconsistency, variation, and missing information.
- **Encoding challenges.** Equally challenging is finding suitable ways to encode all the qualities and quantities relevant to synthesis—all the dimensions of heterogeneity, for example. Each may open a pandora's box of complexities related to faithfully recording the core data, inconsistency, variation, and missing values. Comprehensively encoding the intervention a population received in a trial is illustrative: The treatment may include multiple drugs, at multiple doses, on multiple schedules, with switching strategies employed at certain time points or given certain conditions. Development of an ontology and schema suitable for such data is itself worthy of a research project.
- **Incomplete evidence.** Necessary data (or dimensions of data) may be nonexistent, or difficult or impossible to obtain. Even when the data *do* exist, without advanced natural language processing or an army of data-entry volunteers, it would simply take too long to encode relevant data. (Systematic reviews in the Cochrane database include detailed data for included trials along a number of dimensions like population, intervention, comparison, sources of bias, etc.—in prose, although it is

systematically presented in printed tables.) With details missing, it may be more difficult to produce a trustworthy and reliable synthesis of data.

- **Advanced knowledge required.** Synthesis—such as a network meta-analysis—demands advanced medical knowledge *and* advanced biostatistics knowledge, often on the part of the *reader* as well as the producer. It is *far* beyond the scope of an undergraduate thesis project to take this challenge on, especially in a generalizable way. However, it is absolutely reasonable that algorithms could be encoded and applied in an automated way to solutions built with the prototype, just as simple statistical calculations can be performed with its machine-readable data.

Communication challenges. Communicating synthesized "conclusions," with all the uncertainty, heterogeneity, missing data, and so forth, is another pandora's box of complexity. Only in the most unequivocal case, where all such questions were resolved, would it be possible to provide an unambiguous answer.

The prototype had to "take in" new findings and display them alongside similar findings, without worry about how they relate to one another. Differences—in source, in quality, in intervention, etc.—can be *shown* rather than *interpreted*. That is sufficient for demonstration. Thus a systematic approach to *accepting* and *presenting* evidence was adopted. However, this approach brings with it risks of different sorts. Editorial decisions about *which* dimensions of a finding to show (e.g. should the information be quality rated, and should that quality rating be shown?) are a source of bias, as are visual design decisions (e.g. how prominent to make the quality information, and how to communicate it). It places the burden of understanding on the user, who must interpret the various dimensions and make meaning of them as a whole. No finding is considered more important or relevant than another; data are unweighted, so that a less reliable finding about a certain drug may appear next to a much more reliable one. The process of "aligning" disparate data for consumption can be called *harmonization*. They may be made to conform in presentation, if not in underlying format.

Translation

Because the prototype deals with data about outcomes useful to researchers, but not necessarily to patients, *translation* is a fundamental principle. It is not enough to take findings from evidence and present them *as is*. For one thing, it is likely that with a high-level issue in mind—"I've been sleeping badly. Will methotrexate help me sleep better?"—a patient might not know which "research outcome" to hunt for, nor where to find it. Translation begins with a layer of UI that orients ordinary people to the high-level domain or concept (sleep) first, and subsequently finds appropriate evidence to display. However, because the prototype is intended to be elastic and systematized, it would be an anti-principle to manually *re-encode* sleep-related findings, and build a bespoke sleep-oriented UI. Instead, the roots of an automated way of this translation is necessary. Similarly, translation in this sense includes replacing jargon with plain language. And using visualizations to translate statistical findings into graphical representations that might be more understandable than the statistic as a number.

This principle guided the creation of a simple controlled vocabulary for retrieving appropriate findings for given high-level concepts. It also led to a text-light UI with as little narrative text as possible; in general, data points stand alone, devoid of significant context. That could make elements harder to understand, but it also means that the UI components are more flexible. For instance, a general relative risk comparison visualization component was developed that shows the risk relative to baseline for *any* outcome. It has a very general design that can accommodate evidence about RA improvement after 6 months for multiple medications compared to placebo, or show the relative risk of a single adverse event. It is optimized for neither—it translates from a relative risk statistic to a visualization, but probably does not translate in an optimally understandable way for either case.

Cognizance of bias and framing effects

A discussion of the many potential sources of bias in the prototype, from the methodologies employed in seeking evidence to encode, to the data schema, to technical decisions and tool choices, to UI and visualization choices, could be a lengthy thesis of its own. Suffice it to say that there are *many* sources of bias in the authorship and development of this thesis project, and that live on in the prototype platform. However, this principle insisted only that bias be explicitly acknowledged when possible, even when it decisions were made that introduced a non-ideal bias.

Objectivity is impossible, so there were explicit decisions to guard against bias in areas as diverse as:

- Evidence (e.g. quality, selection, inclusion, omission, and level of detail encoded).
- Balance of gist and precision.
- Visual design (e.g. color, scale, text vs. image, ordering and position of elements).
- Cultural references.
- Linguistic choices (e.g. grouping certain outcomes under a plain language term like “overall improvement”—a deliberate, biasing positive framing).
- Use of data (e.g. inclusion, omission, numbers and charts vs. narrative, dealing with missing or unavailable data).

There are many ways in which end-user interpretation of health information can be biased by design decisions. The way in which a finding is presented—if it is “framed” positively or negatively, made larger, colored red or green, displayed visually or as a number, expressed as an absolute value instead or as relative odds, and so on—can introduce a *framing effect*, a cognitive bias that tends to lead understanding one way or another. All design decisions are editorial in this way; an understanding of known best practices from risk communication literature (Akl et al., 2011; Perneger & Agoritsas, 2011; For example Trevena et al., 2013), as well as general design training (e.g. in gestalt psychology) can help during the intuitive design process.

Reinforcing the principles of systematization, elasticity, re-usability, and the notion that the prototype is a general system, essential design decisions

were made to both adopt certain best practices, and to make an effort to avoid some framing effects (while accepting the introduction of others). For example, the *Navigator* application prototype is a greyscale UI, and has little framing text. Both of these decisions are intended to "let the data" speak without introduction of color or significant persuasive rhetorical language as potentially biasing elements. However, those choices themselves introduce a potential framing effect—the attempt to be "neutral" and to focus user attention on the data (findings) themselves may bias how meaning is made. Study of the prototype and its present visualizations with real patients and clinicians would uncover some of these effects. On the other hand, much literature (including articles referenced above) have pointed to the less-biasing effect of presenting health-related outcomes as absolute risks, and using icon arrays (Galesic, Garcia-Retamero, & Gigerenzer, 2009) to help low-numeracy people understand such data, and so the *Navigator* opts to show those data more often than relative risks. The first re-usable UI components are those that take into account some of the best practices from that risk communication literature.

The best way to summarize the state of bias and framing effects in the prototype platform and example applications like the *Navigator* is that there *are* such issues. One immediately obvious example is that it presently contains a tiny sliver of evidence about RA medications, but does not (and can not) clearly communicate the boundaries of that sliver of evidence in a quantitative way. A layperson could easily be fooled into thinking that data they see in the *Navigator* are complete and authoritative, when they are not. Ultimately, each tool built with such a platform should do its best, given its purpose, to address bias. The next step for uncovering the details about bias and framing effects in the prototype would be research and discussion with outside parties—specialists, end-users, and so on.

Tailorability

Health information and guidance tailored to its audience, and sometimes entirely personalized to "speak to" its audience as an individual, is often more effective than a generic equivalent (Hawkins, Kreuter, Resnicow, Fishbein, & Dijkstra, 2008; Kreuter, Oswald, Bull, & Clark, 2000). Here, tailoring might be about the user interace itself (how the evidence is presented), and also about highlighting evidence that is (either directly or in an inferred way) applicable to the end user. The capacity to so personalize the prototype—to as many dimensions as possible, such as age, sex, gender, lifestyle, values, health history, sociocultural setting, literacy level, numeracy level, and physical capacities—was a desired quality, but difficult to support. While this principle was considered early on in development, supporting it with appropriate atomic data encoding was deliberately avoided to focus on other problems. Because population details—for example, demographic and health status data about participants in trials—are reported so heterogeneously, and often do not resolve to an easily encodable pattern or conclusion (like *mostly women with severe active RA of > 5 years duration*), it was not clear that the additional difficulty that finding a suitable way to capture those data would "pay off" for the purposes of the prototype. However, the prototype *does* allow some tailoring in terms of filtering: For example, the medication list can be limited to those that are suitable for a basic set of preferences around dosage form and whether one drinks alcohol regularly. On the other hand, little effort was made to tailor *messaging* or the presentation of findings themselves.

Working with evidence

There is a dizzying array of sources of potentially valuable evidence about RA medications. This project was largely inspired by the lack of available numerical, machine-readable data from biomedical research with which to create rich, interactive create data-driven resources. Thus sources from the scholarly literature were prioritized. Other kinds of evidence—patient-reported data from Web sites like PatientsLikeMe (see Chapter 2), or interviews from sources like Healthtalk.org (also discussed in Chapter 2)—could complement the scholarly sources quite nicely in an online resource were it possible to incorporate them.

Outcome data from systematic reviews or meta-analyses was preferred over findings from individual clinical trials, product literature, observational studies, and other sources—although there are data from such sources in the spreadsheet. Despite the name *systematic* review, not all findings or results in systematic reviews are *reported* as systematically as they are reviewed. One review of conventional DMARDs' effects on pain in inflammatory arthritis (Steiman et al., 2013) illustrates the challenge of working with such sources. In some cases, systematic funnel plots and tables were shown for interventions on a per-disease basis, and in other cases the data were selectively reported in sentence form. Systematic reviews and meta-analyses tended to report on included trials' quality, assessing them for randomization, blinding, and other sources of bias. Many (especially Cochrane Collaboration reviews) employed the GRADE estimate-rating system (G. H. Guyatt et al., 2008). Even still, some reviews and meta-analyses were of concern because of author links to pharmaceutical companies (For example Jansen, Buckley, Dejonckheere, & Ogale, 2014) although these links were usually disclosed. Some systematic reviews, and especially network meta-analyses employed complex biostatistical methods that were beyond the comprehension of a beginner in this domain, casting doubt on whether it was appropriate to encode findings from them. For the practical purpose of finding data for the prototype, descriptions of many of the more complex methods had to be glossed over. Ultimately, little data were encoded from such sources, illuminating a weakness of "independent study" of this domain. "On-the-ground" collaborators with specialized knowledge would be necessary to build on the prototype thus far.

Selection of evidence

In the event that a "real" platform were to grow from this prototype, the review and selection of evidence would become a much larger methodological concern. For the purposes of the prototype, evidence about RA medications was selected according to these heuristics, always with the guiding principle of sufficiency in mind:

- Is it specific (about a specific DMARD, rather than a group of medications)?
- Is the population general enough for the data to be relatively widely applicable?

- Is it relevant to a domain of health of concern to RA patients? For the moment, this heuristic meant that findings about erythrocyte sedimentation rate were excluded, for instance.
- Is the finding relatively easily encodable? In a few cases, findings are too complex to encode—but that is rare.
- Is the finding high quality? This heuristic required a good deal of reading, since the humble author of this thesis project is a newcomer to the world of medical evidence. However, in general findings that came from sources with lower risk of bias were preferred.
- Are the heterogeneity and uncertainty in the finding acceptable? (See below for an elaboration.)

Statistical literacy

There are simple heuristics to evaluate the quality of sources of evidence—for instance, preferring peer-reviewed meta-analyses and systematic reviews whose authors have few or no ties to the pharmaceutical industry. However, these heuristics are ultimately inadequate, both for a first-pass selection of evidence and for the future work of aggregating and synthesizing data. Statistical methods—such as those employed in the production of network meta-analyses—are so specialized that there exists a scholarly discipline—biostatistics—devoted to them. And multiple academic journals are devoted to the topic. For the practical purpose of finding data for the prototype, descriptions of methods had to be glossed over. There is a danger in assuming that the statistical methods are appropriate, and therefore the estimates of effect trustworthy.

Heterogeneity and uncertainty

These two concepts suffused selection and use of evidence in the prototype. In an ideal world, data would be uniform—that is, *what* is measured, *how* it is measured, and its encoding and documentation would be identical from source to source. However, that is obviously not the case. Systematic reviews and meta-analyses bring a sense of uniformity to evidence by applying advanced statistical methods to data. A good deal of the prose in these scholarly publications is devoted to explaining heterogeneity and defining more uniform concepts to organize the data in the review. For example, consider the following excerpt under the heading "Evidence base" from Jansen et al. (2014), a systematic review of the comparative efficacy of biologic DMARDs on PROs in patients with an inadequate response to conventional DMARDs. **Bold** emphasis was applied to highlight statements addressing heterogeneity.

> Most of the trials were multi-centred and included patients predominantly from Europe and North America. The RCTs were generally considered to be good quality **(Jadad score range 3–5)**. All included trials were double blind with appropriate description of drop out of subjects, **although the method of randomisation and blinding was not always reported**. The **majority of the studies included adult patients with diagnosis of RA based on the ACR 1987 revised classification criteria.** All studies included DMARD-IR patients. **Although the definition of DMARD-IR** [intolerant or inadequate response] **varied somewhat between the studies,** it was

> most commonly defined as patients with active disease despite of [sic] previous treatment with traditional DMARDs. **The traditional DMARD was often specified to be MTX** [methotrexate]**, although in fewer studies it was unspecified.** Other definitions included inadequate response to prior DMARDs, or patients who are either intolerant to MTX, or the use of MTX is inappropriate. The TEMPO trial included patients who were non-responders to DMARDs but disqualified patients who had failed MTX treatment. Given this difference, the study was excluded from the network meta-analysis. **The definitions of active disease varied in terms of the minimum levels of ESR** (10 mm/h, 28 mm/h) and CRP (2 mg/dl, 1 mg/dl, 1.5 mg/dl, 7 mg/ml), **as well as in terms of the minimum number of required tender and swollen joints. Not all studies reported whether RA disease duration and DMARD treatment duration determined eligibility.**
>
> (Jansen et al., 2014)

Here is another example from Wee et al. (2012), a systematic review of the effect of biologic DMARDs on work participation, under the heading of "Statistical analysis:"

> A meta-analysis to assess the overall effect of biological agents on work participation in RA could not be performed due to extended heterogeneity with respect to study populations, outcome measures and statistical analysis. Therefore, narrative summaries are provided.
>
> (Wee, Lems, Usan, Gulpen, & Boonen, 2012)

Heterogeneity in outcomes, measurement, populations, intervention details, follow-up time, study design, etc. all contribute to the difficulty of comparing, synthesizing, harmonizing, or even just "bringing together" findings from multiple sources. From the same systematic review is this note about heterogeneous outcomes:

> Almost all studies used different approaches to assess work outcome comprising self-composed questionnaires, validated instruments or existing databases reporting on work outcomes. Although 24 instruments are available to assess absence from work and/or presenteeism, only few a studies used one of these instruments (work productivity and activity impairment questionnaire, RA work impact scale, work productivity survey–RA, work limitations questionnaire, workability index and health and labour questionnaire). Even then, comparability remains limited as these instruments also differ in recall, concepts of absenteeism and presenteeism and whether or not impact on work should be attributed to RA or overall health. Moreover, some studies presented results on a group level (means/median) while others on the individual patient level (proportion of patients with sick leave or presenteeism). Also, most studies used different definitions to describe employment status such as: being employed, additional years worked, work disability rate (official or self-perceived) and job loss. For example, one study defined employment status as differences between groups in gaining or remaining employed. Another study reported on

> hours working per week, which made it unclear whether this referred to employment status (contract hours) or also included productivity loss due to sick leave.
>
> (Wee et al., 2012)

These situations are enough to make one throw up one's hands in exasperation and defeat, especially when there is otherwise little evidence about the domain of interest—for example, RA treatments and work-life outcomes. In some cases statistical methods exist for pooling and estimating effects from studies that use different measures, but these methods require substantial expertise and thus could not be applied in an automatic way. In many cases, heterogeneity (especially in terms of the population that was studied) was simply accepted—however, it *severely* limits the applicability of data in the prototype to decision making. A candidate for future work is a rich and applicable encoding system for the population to which findings apply, so that users could tailor presentation to only include evidence that is more likely to be relevant to them.

With respect to uncertainty, there is no best place to begin. In many sources, findings were noted as not being statistically significant—especially comparing medications to one another or to placebo. That note was never recorded, although from other data in the schema (population sizes, confidence intervals, etc.) rough approximations of statistical uncertainty could be recreated in software. For the purposes of the *Navigator* prototype, even those confidence intervals have been omitted from visualizations. This is a serious weakness of the prototype and an area for future work—especially in terms of visualizing uncertainty and studying how understandable those visualizations are (Gresh, Deleris, & Gasparini, 2012; Roth, 2012; Spiegelhalter, Pearson, & Short, 2011).

Encoding and data schema

Findings were extracted by reading source articles and manually typing them into a spreadsheet. They were encoded in a schema which emerged during the course of development. Findings were not encoded systematically: Although the encoding strategy accommodates many kinds of outcomes, and therefore many kinds of data, the focus was on data points that would demonstrate practical value for some of the top issues RA patients might be concerned with, described in the next section. For any source, a limited high-level set of findings were extracted and encoded. Those are the data that power the *Navigator* prototype and are discussed in more detail in the next section. When findings *were* encoded, they were encoded as completely as possible according to fields in the schema.

Although the schema for outcomes itself is not fixed (it is constantly changing as the prototype continues to develop), as of February 2016 it consisted of the following fields, all visible in the Google Spreadsheet referenced earlier in this chapter:

> **which** *intervention* | *comparison* | *population*
> Which of an *intervention*, *comparison* (comparator), or *population* the line of the spreadsheet (finding) is about. For any case where there is an intervention versus a comparator, there will be two lines in the spreadsheet for that comparison: One where the **which** is *comparison* (the baseline or assumed risk, or comparator data point) and one where

the **which** is intervention (the intervention data point as compared to the comparator case).

measure
Which outcome this finding is about. Each time a new outcome was encountered, it was added to the *Measures* spreadsheet. Examples include *acr_50* (ACR50 response) and *serious_ae* (serious adverse event).

measure_detail
Additional details about the measure. For instance, for *ae* (adverse event or side effect), this field has the name of the adverse event (e.g. *headache*).

metric
The statistic used to record the outcome. Each time a new way was encountered, it was added to the *Metrics* spreadsheet. Examples include *ar_100* (absolute risk or frequency out of 100), *rr* (relative risk), *mean_score* (mean score), and *count*.

value
The numerical value, expressed as expected by the metric. For example, *24* is a suitable value for *ar_100* (meaning 24 out of 100 or 24%).

value_ci_low
The lower bound of the 95% confidence interval. Only 95% CI is supported; it would have been possible to support other measures of confidence and to include a field in order to indicate what measure was being used (e.g. interquartile range or 95% CI) but it was not deemed relevant for patient exploration of evidence—yet. This is expressed in the same metric as **value.**

value_ci_high
The upper bound of the 95% CI.

grade
The GRADE level of evidence for this finding, if available. This probably should be renamed **quality** or **certainty** and redesigned to include a field indicating what system is used.

n
The number of participants in the study or studies, if applicable.

n_type
If only the overall **n** is available (i.e. for all the participants who participated in a series of studies), this value is *total*. If the **n** refers to the number of participants who were in a particular group (e.g. the intervention group or the comparison group), this value is the same as in **which** (e.g. *intervention*).

duration_low
The earliest or only follow-up time, expressed as an integer.

duration_high
The latest or only follow-up time, expressed as an integer.

duration_interval
The period (i.e. unit of time) that the duration is expressed in. For example, *day, week,* or *month*.

population
A description of the population that was studied, if appropriate. This is not broken into machine readable parts, because it was not deemed necessary for the prototype—yet.

intervention
A comma-separated list of therapies that made up the intervention. In general, generic medication names are used. The primary intervention is listed first. The generic name is used by the data processing code to look up basic medication information, perform filtering, etc. Examples include *methotrexate* or *tocilizumab,dmard*.

comparison (comparator)
A comma-separated list of therapies that made up the comparator. If the finding is not about an intervention on its own, devoid of comparison, this is blank. If the finding did involve a comparison (as is the case when a relative risk was reported), the comparator is listed just as **intervention** would be. Examples include *placebo* or *placebo,methotrexate*.

dosage
If **which** is *comparison,* the dosage of the comparator, if appropriate. If **which** is *intervention,* the dosage of the intervention. Dosage if usually expressed as a number, then a space, then a unit. Dosage is compressed into a single string here, since it can relatively easily be parsed if necessary. The individual parts of the dosage string being encoded separately did not seem relevant to the high level exploration of evidence by RA patients. Examples include *8 mg/kg* and *25 mg*.

dosage_form
Like **dosage**, refers either to the *comparison* or *intervention*. This is the dosage form or route of administration. Any string of text is valid here; in general the values are *oral*, *intravenous,* or *subcutaneous*.

dosage_frequency
Like **dosage**, refers either to the *comparison* or *intervention*. This is how many times the comparator or intervention was taken per dosage interval. The interval is calculated by the next two fields.

dosage_multiple
This is how *many* **dosage_interval** periods elapse per **dosage_frequency.** For example, if this is *4* and **dosage_interval** is *week*, that means “every 4 weeks.”

dosage_interval
The period (i.e. unit of time) that the dosage interval is expressed in. For example, *day, week,* or *month*.

source
A hyperlink to the source from which this finding was extracted or in which it was found.

notes
Any notes or other important annotations about this finding.

kind
The kind of source that this finding came from. This is not strictly enforced, but could be just as *metric* and *measure* are. Values include *systematic review, randomized trial*, and *meta-analysis*.

What people want to know about RA treatments

There is evidence about patient preferences for information about RA treatments. In general, people with RA prefer to be fully informed about treatment options and alternatives (L. Fraenkel, Bogardus, Concato, & Felson, 2001). However, specific outcomes that patients care about are incompletely studied (domains of interest are known, but measures are still in development) and relatively infrequently reported. In a survey of 254 RA patients in the U.K. (99% of them White) in which respondents rated the importance of plain-language outcomes, the top 6 were: "Less pain, doing everyday things, no more (visible) joint damage, more mobility, enjoy life, and more independent" (Sanderson, Morris, Calnan, Richards, & Hewlett, 2010). Pain, ability to perform everyday activities, fatigue, psychological distress, ability to cope, overall well-being, sleep, and work and social life are all important outcomes to RA patients (Gossec, Dougados, & Dixon, 2015). When weighing benefits and risks of treatment options, they may value reduction in the risk of bothersome (e.g. diarrhea) and grave (e.g. serious infection) side effects *more* than the likelihood of feeling better (L. Fraenkel, Bogardus, Concato, Felson, & Wittink, 2004), although benefit (improvement in RA) is still of high utility. The RA Choice decision aid issues—onset, cost, side effects, route of administration, and other considerations like pregnancy, alcohol consumption, and concomitant tuberculosis risk—also underscore kinds of information valuable to both patients and clinicians in decision making (Barton et al., 2014). What people want to know depends in part on awareness of what there is to know, which tools like the RA Choice DA can help make explicit.

What the research community produces

Research on treatments for RA is produced by and for the biomedical community. OMERACT, a research collaboration on clinical trial outcome measures in rheumatology, formed in 1992 and has over the past two decades performed research in and promulgated standardized outcome measures in RA clinical trials (Boers et al., 2014). The most well-known and commonly used

measures are the "core set" that was adopted (published) by the American College of Rheumatology (ACR) in 1993 (Felson et al., 1993):

1. **Tender joint count (TJC).** Assessed by a physician by physically manipulating and examining 68 joints and classifying joints as either tender or not.
2. **Swollen joint count (SJC).** Assessed by a physician by physically examining 66 joints as either swollen or not.
3. **Patient's assessment of pain.** Assessed by the patient, without mediation by a clinician (therefore a PRO), using an instrument like a 10 cm visual analog scale, Likert-type scale, or similar. The question asked and scale anchors vary.
4. **Patient's global assessment of disease activity.** Assessed by the patient, without mediation by a clinician (therefore a PRO), using a 10 cm visual analog scale, Likert-type scale, or similar. The question asked and scale anchors vary.
5. **Physician's global assessment of disease activity.** Assessed by a physician, using a 10 cm visual analog scale, Likert-type scale, or similar. The question asked and scale anchors vary.
6. **Patient's assessment of physical function.** Assessed by the patient, without mediation by a clinician (therefore a PRO), using a validated instrument. There are many potential instruments or measures; the heterogeneity in these measures introduces harmonization challenges.
7. **Acute-phase reactant value.** A laboratory assay, either erythrocyte sedimentation rate or C-reactive protein level, both ways of measuring systemic inflammation.

That these measures exist and are standardized is extremely helpful for comparing treatments to one another. These outcomes are used extensively in clinical trials research and in systematic reviews and meta-analyses of RA treatments. Changes (improvement) in these measures form the base of evidence underlying RA clinical practice guidelines and evidence-based practice. *Improvement* is usually measured using a dichotomous outcome measure based on the core set. *ACR20*, for example, means that someone showed improvement of at least 20% in both TJC and SJC, as well as in 3 of 5 additional measures. However, there remains the question of whether these measures make sense to patients—especially when someone has a new diagnosis, how might they understand the meaning of an improvement in tender joint count? Would they understand what ACR20 *feels like?* Would they understand the meaning of a 2 cm reduction in their "global assessment of disease activity," if they started at the 7 cm point on a scale? There is an understanding among the research community that patients have concerns that might not be addressed specifically by these measures, and that development of a "patient core set" to complement the research-oriented outcomes may be valuable (J. R. Kirwan et al., 2009).

Incorporating patient-reported outcomes in research about RA medications has been a slow process relative to the pace of change in technology—though that is understandable given the rigor with which such outcomes must be researched and developed. At the OMERACT 5 meeting in 2002, a research agenda on patient-oriented measures was described (J. Kirwan et al., 2003),

but as of 2015 PROs in many important domains remained more or less absent from clinical trials on RA medications (Gossec et al., 2015). In the intervening time, smart phones had become ubiquitous along with high-speed wireless Internet access, and so to almost *all* data-driven online resources that are commonly used today by the general public (e.g. GPS navigation applications, travel booking tools, etc.). With or without ready-made PROs, "the challenge remaining for rheumatologists is how to effectively communicate the risks and benefits related to the many options now available for the treatment of RA" (L. Fraenkel et al., 2001).

Can whether people *thrive* be measured?

Given what matters to RA patients, finding ways of communicating evidence about treatment options *on their terms* should be a priority. Directly applicable PROs are ideal, and they might help answer the question of whether an intervention will help them thrive, and in what ways. It is naïve to think that a clear answer would magically emerge from PROs; it is nevertheless a legitimate aspiration to want to be able to frame exploration of evidence in terms that are as closely aligned as possible to the mental models or ways of thinking and concern that people adopt when learning about treatment options and engaging in medical decision-making. There are common scenarios in which the evidence in the scholarly literature that *is* available may be difficult to communicate:

- **There is relatively little evidence on an outcome that matters to patients.** For example, on work and social life, or sleep.
- **There is a good deal of evidence but it is recorded using measures that *indirectly* bear on health domains or concepts that matter to patients.** For example, ACR50, which means 50% improvement in TJC and SJC as well as on 3 of the 5 additional ACR core set outcomes. It is difficult to map or translate that outcome onto patient-important domains (e.g. pain, sleep, work, etc.). Adverse events are another illustrative example. Clinically-oriented evidence, like systematic reviews, frequently report the proportion of people who withdrew from clinical trials of a medication due to a serious adverse event. That is an important measure of medication *safety*, but does not necessarily address the fact that patients consider non-serious side effects like nausea and diarrhea as important as serious side effects.
- **PROs are used, but data are subsumed by composite measures.** For example, someone "experiencing" ACR50 (i.e. 50% improvement) probably experiences a reduction in pain, but not necessarily a reduction that they would report as a 50% improvement on a 10 cm visual analog scale. The data on pain reduction itself are obfuscated. Even still, there is the legitimate question of whether someone who is not used to "recording" their own experience of pain would even understand the data if it *were* available.

In these cases, it is probably important to *translate* findings into a more understandable form. Since it is more or less impossible to explain *precisely* what happens to the whole physical, psychological, and social experience of someone who experiences 50% improvement, translation is certainly difficult. There are some basic techniques that might be applied:

- **Summarize in prose.** For example, *after 1 year of treatment with Enbrel (etanercept), it is estimated that about half the people taking it would experience a 50% improvement in their RA–tender and swollen joints, pain, and ability to do their daily activities.*
- **Use visualizations.** These could include icon arrays, line charts, or other graphics.
- **Complement with other evidence about what people who have experienced this outcome may also experience.** For example, *people who had a 50% improvement in their RA after 6 months were doing pretty well in terms of their ability to keep working 5 years later. On average, they only had to take about 4 days off work per year.*
- **Complement with anecdotes from people who have experienced this outcome.** Interviews–in text, audio, or video formats–with people who "just" experienced the outcome (e.g. ACR50–50% improvement–after 6 months or 1 year) could enrich understanding of the measure.

Of course, there is also the problem that these data are not commonly available in machine-readable form, so they cannot even be easily integrated into an automated system for translating evidence.

Settling for an *issue* focus

To facilitate exploration of evidence in the prototype, and to achieve a modicum of automated "translation" of extant data, an effort was made to support *retrieval* of evidence in terms of *issues.* Like the Mayo Clinic decision aids (see Chapter 2 for references), RA Choice decision aid (Barton et al., 2014), and MAGIC decision aids (Agoritsas et al., 2015), the *Navigator* prototype organizes evidence presentation according to these topics. Put simply, an issue is a patient-important question or concern. There is theoretically no conceptual boundary around how vague or abstract an issue can be; for practical purposes, it makes sense to consider issues as at the level of "exercise" or "feeling well" or "daily routine" as the most abstract, and "pain" as the lowest-level, or closest to atomic clinical measures.

The prototype platform supports arbitrarily tagging outcome measures with plain language terms that reflect these issues. The *Navigator* prototype UI demonstrates use of just a few of these *tags* to help people understand the impact of RA medications in terms of a few of those important areas:

- **Basic issues.** Practical concerns like medication names, cost, route of administration, contraindications, etc.
- **Overall improvement.** Reduction in the signs and symptoms of RA. In other words, the *benefits* of treatment on the disease process and its symptoms.
- **Pain.** Reduction in pain.
- **Work.** Ability to continue working.
- **Side effects.** Both serious and non-serious adverse events associated with medications.

Subsequent discussion focuses on the practical problems of working with and designing for evidence used in the *Navigator* prototype, though it applies more broadly. Before diving into those five issues, it is necessary to explain how evidence is "translated."

Harmonization & translation layer

Because there is sometimes a disparity between what people want to know about RA medications and the outcomes found in research data, it is not enough to just encode findings. Imagine that you have just been diagnosed RA, and have a question about how well you will generally feel if you take methotrexate by injection, once a week. "How well you feel" is a subjective, complex evaluation of your experience, but in this case it can reasonably be reduced to a composite of how the medication might make you feel, how it might affect your RA, and how that change in your RA is likely to affect how you feel and consequently how your life might change. If you feel better, and the medication is tolerable, perhaps you will feel a *lot* better than you do now, partly because of the relief you will experience from the medication but also because you might exercise more, eat differently, and so forth. Although no evidence can answer the question fully nor predict the likelihood of methotrexate working, there are sources that address some of the questions individually:

- **How the medication might make you feel.** Data about adverse events, testimonials or other anecdotal evidence from patients who have taken methotrexate, the advice of a physician based on clinical experience with hundreds or thousands of patients, etc.
- **How it might affect your RA.** Data from clinical trials or systematic reviews, clinical practice guidelines, epidemiological studies, testimonials or other anecdotal evidence from patients who have taken methotrexate, physician knowledge, etc.
- **How the change in your RA is likely to affect you how feel.** Epidemiological studies, clinical trials, cohort analyses from patient registries, testimonials or other anecdotal evidence from patients who have taken methotrexate, physician knowledge, etc.

For each rough category (such as data about adverse events) there are many potential sub-sources. In that case, they may include product literature, post-marketing safety surveillance databases like openFDA (Kass-Hout et al., 2015), various kinds of retrospective studies, clinical trials, systematic reviews, and even anecdote. Each source may have—in fact is *likely* to have—its own way of encoding findings. For example, there may be detailed data about how likely

a particular side effect is, or there might be a single number describing how many people in a given population stopped therapy because of any adverse event. The same general category of adverse event (say, gastrointestinal distress of some sort) may described using any number of idiosyncratic terms. Perhaps a given adverse event was more common when methotrexate was combined with another DMARD in a common therapeutic scenario. The challenge of *harmonization* is finding a conceptually and technically adequate way of aligning these data points, such that they can be presented sensibly to someone with a question like "how well will I generally feel if I take methotrexate by injection, once a week?" Data-oriented harmonization might focus on ensuring maximum quality and consistency in the data. User-centered or patient-centered harmonization might focus on abstracting away unnecessary nuance to provide useful conclusions. The practice of medicine is messy, and with imperfect evidence decisions must still be made.

The prototype as originally envisioned would feature the capacity to harmonize (or make comparable) similar findings measured with slightly different outcomes. Then, it would use data visualization or interactive UI to make the data more understandable than if it were presented in a table or paragraph of text. Because of time constraints, lack of expertise, and a combination of homogeneity in certain outcomes and lack of evidence for others, only a very limited capability has been built so far. It effectively has two parts:

- A controlled vocabulary to "organize" findings and guide user navigation.
- Data visualizations (also guided partially by the controlled vocabulary) to take arbitrary outcome data and visualize them.

Figure 3-3 shows how the prototype "retrieves" evidence according to the user's request, given a top-level *issue* of concern like *overall improvement*. Data processing code and individual data visualizations employ the controlled vocabulary to find the appropriate data, process it, and display it in the desired form. In the future, qualitative data might be an effective complement alongside individual findings. For example, video interviews with patients describing their experience of pain relief could be retrieved alongside information about how well medications may reduce pain.

Controlled vocabulary

Controlled vocabularies—like that used by the Library of Congress to index publications—are sets of predefined terms, and sometimes relationships between them, that may be used to organize or describe concepts in a given domain and facilitate information retrieval. They are routinely used in medicine (Cimino, 1998). A simple controlled vocabulary for the prototype grew organically as findings were added. (The controlled vocabulary is also part of the data spreadsheet.) It is disentangled from data, such that someone reviewing evidence and adding findings to the data spreadsheet does not have to be at all concerned with translating to patient-friendly terms. The translation layer of this vocabulary and UI and data visualizations "takes care" of that translation.

- **Measures.** These are the various ways that effects of RA medications are gauged or coded—generally *outcomes*. For example, in the prototype *ACR 50* is considered a *measure* (though ACR 50 is itself a composite measure),

and so is *erythrocyte sedimentation rate*. Typically, these measures are the key kinds of findings that are reported in research about RA medications. Ordinarily they are reported in jargon, and represent investigations of outcomes or factors important to science rather than necessarily to issues that patients are likely to understand or care about. Each data point in the spreadsheet reports a finding for one of these measures.

- **Metrics.** A single statistic, recording a finding in a given measure, has to be reported in a certain way. For example, if the outcome was people who achieved the *ACR 50* level of arthritis improvement, how was this improvement measured? The relative chance someone might achieve *ACR 50* with one treatment compared to another? A count of people who achieved it? The absolute frequency—out of 100, or perhaps 1000—of people who achieved it? These are the details described by the *metric*.

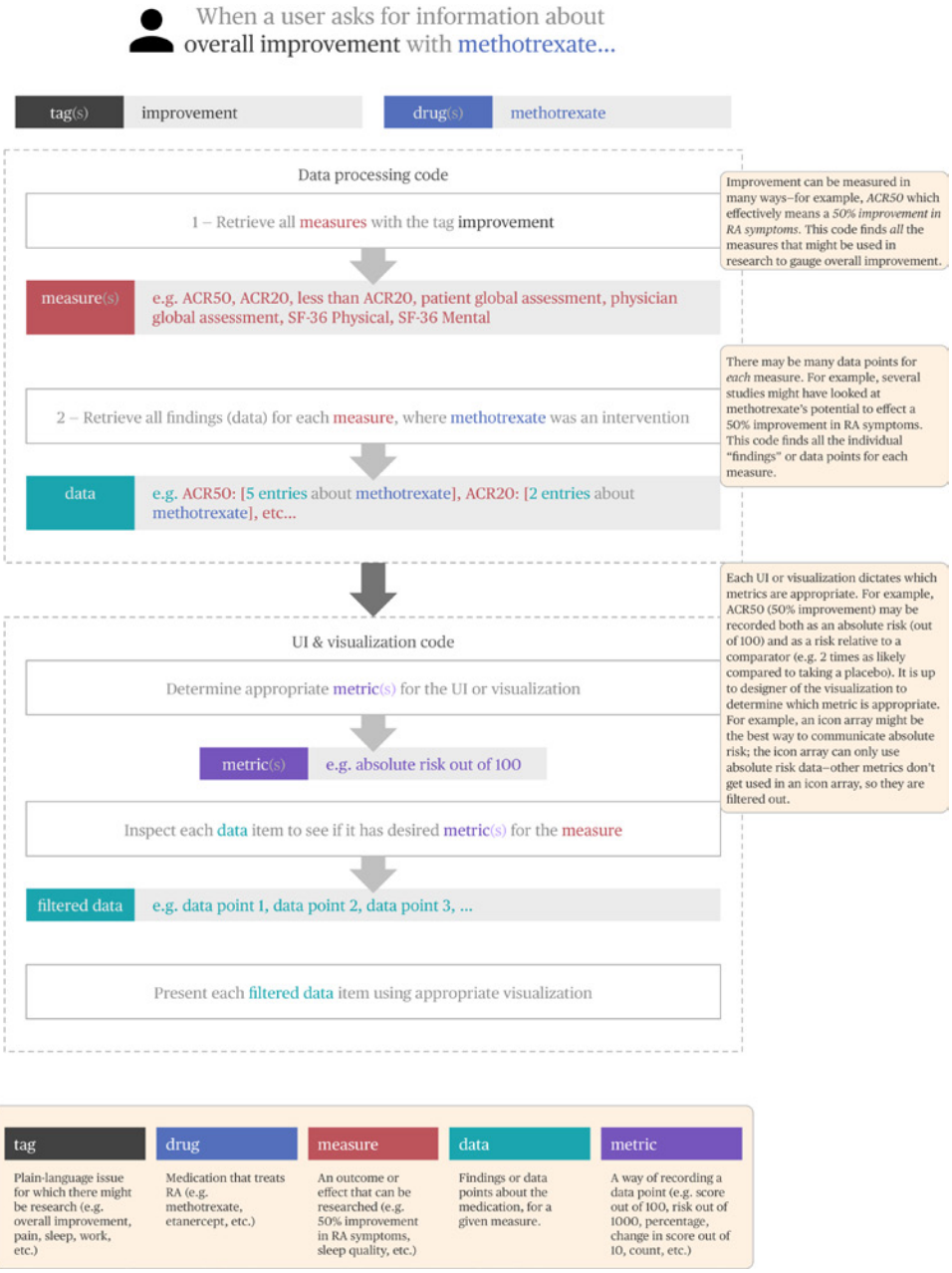


Figure 3-3
Schematic diagram of the prototype's simple controlled vocabulary.

Each data point in the spreadsheet reports a finding using one of these metrics.

- **Tags.** Because measures are usually researcher-oriented, even a well-visualized representation of a particular statistic might mean nothing to an ordinary patient. After all, one might ask "what does *ACR 50 mean?*" A succinct description helps, but is insufficient to provide a kind of wayfinding guidance given a potentially confusing array of measures and apparently endless research. In order to provide straightforward navigation *to* research that may bear on patient-important issues, measures are *tagged* with plain language terms that describe what the measure is about. For example, *ACR 50* is primarily about overall improvement, and therefore tagged as such. *TJC* (tender joint count) is mostly about pain (as in painful joints), so it is tagged as such. These "tags" are used by the data processing code to find and group data.

The controlled vocabulary is ultimately responsible for finding the evidence relevant to a given high-level *issue* like pain.

The *Navigator* prototype: Exploring evidence about 5 issues related to RA medication options

Basic issues

As described in the introduction, living with RA is complicated. Medical treatment is an overarching concern in that its effects suffuse the experience of living with RA, but *contrasted* with living with RA, deciding on a medication is a relatively small effort. However, the medication itself *does* demand tangible effort and attention. The first category of information in the *Navigator* prototype deals with these tangible concerns and basic facts about the medications–many of which (e.g. cost, route of administration, and concerns like pregnancy and alcohol consumption) emerged from the user-centered research process employed to develop the RA Choice decision aid (Barton et al., 2014). They include:

- **Medication name and brand names.** It may be important to know that drugs go by multiple names, and that material that someone encounters could refer to the medication by any of those names.

- **Drug class or type.** People may hear about "biologics" or "TNF alpha inhibitors," so medications that fall into this category should be clearly marked as such.
- **Route and schedule of administration.** People may have preferences for one form over another; for example, someone might refuse to perform a self-injection. Others may not feel comfortable with their capacity to take two pills a day, and still others may not want to travel to a clinic for infusions.
- **Cost.** Although ultimate out-of-pocket cost varies with insurance coverage, a rough idea of the cost may be a critical concern for patients. Some of them can cost upwards of $40,000 a year.
- **Lifestyle adjustments, comorbidity risks, and contraindications.** Although a physician would obviously know and discuss these concerns with their patients, patients may want to know which medications are safer given their personal health status and lifestyle considerations. For example, someone who drinks alcohol regularly and does not want to (or cannot) give that up should be able to learn which medications are safer for them. That is equally true for people with comorbid conditions.

This information satisfies fundamental requirements around informing people about treatment options. It also may play a role in someone anticipating whether they have the capacity to integrate a treatment into their life, helping to find the "right care" that is minimally disruptive (Leppin et al., 2015). The data supplying this part of the platform largely came from the RA Choice decision aid (Barton et al., 2014) and a systematic review of use of biologic DMARDs in patients with RA and comorbid conditions (Richards et al., 2015). Additional evidence that might enrich this domain includes other lifestyle impacts (e.g. on eating and food), whether side effects are tied to the schedule of administration, and necessary clinical visits or routine tests associated with the medication.

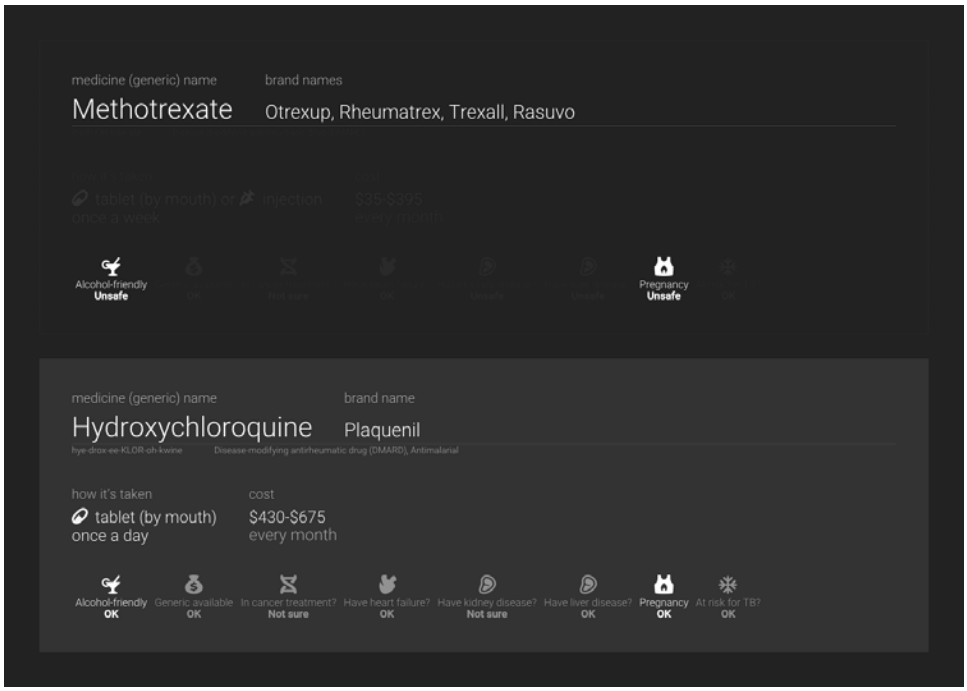


Figure 3-4
Basic issues visualization in the *Navigator* prototype.

Because medications are the primary "object" of concern in the prototype, this basic medication information is encoded as a JavaScript object in the source code, instead of in a spreadsheet like other evidence.

Presentation

Each medication is represented in its own rectangle—a "card" of sorts (see Figure 3-4). The generic drug name is largest, and common brand names listed beside that. Drug names can be difficult to pronounce, so a phonetic transliteration sits beneath the generic drug name. The route and typical schedule of administration are relatively large, and situated next to typical cost. The bottom of the card has a "risk strip" of icons and titles corresponding to lifestyle considerations, comorbidities, and risk concerns. For each, the medication is labeled *OK* if it is safer, *Unsafe* if it is contraindicated, or *Not sure* if no evidence was encoded or it is unclear.

The cards respond immediately to filtering. The user of the *Navigator* prototype can tap or click on preferred dosage forms or preferences, such as a checkbox called "Safer for liver disease." Medications that do not match the selected preferences are reduced in contrast. In the "risk strip," each that conflicts with the user's preferences is boosted in contrast so that the "reason" why a medication was disabled is expressed visually.

Overall improvement

This issue attempts to capture and present evidence about the benefit of medications (i.e. positive outcomes related to controlling RA and its signs and symptoms). For reasons already discussed, this is a particularly tough domain. One might expect patient questions like "will this medication help me feel better?" But it is hard to answer that question given the ways that RA medication effectiveness is measured in research. The ACR core set is the gold standard of measuring benefit outcomes. ACR20 (20% improvement) is considered sufficient for a drug company to claim a "reduction in the signs and symptoms of RA" (Food and Drug Administration, 1999), but it is not clear what that *feels like* from a patient perspective. ACR50 (50% improvement) and ACR70 (70%) improvement are routinely reported. Other measures are also found in the literature, including several ways of measuring remission (absence of signs and symptoms of RA disease activity), the component measures or instruments of the ACR core set (e.g. disease activity score, health assessment questionnaire scores, etc.).

Working with evidence about overall improvement

For most of the 12 DMARDs in the prototype, there was evidence available in systematic reviews from the Cochrane Collaboration library (Katchamart, Trudeau, Phumethum, & Bombardier, 2010; Lethaby et al., 2013; Lopez-Olivo et al., 2014; Lopez-Olivo, Amezaga Urruela, McGahan, Pollono, & Suarez-Almazor, 2015; Maxwell & Singh, 2009; Mertens & Singh, 2008; Navarro-Sarabia, Ariza-Ariza, Hernandez-Cruz, & Villanueva, 2005; Osiri et al., 2003; Singh et al., 2009; Singh, Beg, & Lopez-Olivo, 2010; Suarez-Almazor et al., 2000). Other

sources were also consulted (Golicki et al., 2012; Jansen et al., 2014; Katchamart et al., 2010; Smolen et al., 1999; Strand et al., 2014; Williams et al., 1985).

As is the case for *all* the findings encoded in the Google spreadsheet, data were manually re-encoded in the schema that emerged during development. They were not systematically and completely encoded. For the purposes of prototyping and according to the guiding principle of sufficiency, only *some* findings were extracted. Since the prototype is intended to *demonstrate* that findings from multiple sources could be shown side by side in the prototype, it was not necessary to extract *all* outcome data. During the development process, enough literature had been reviewed and that ACR50 outcomes as reasonable for prototyping purposes. Other outcome data were sometimes extracted and can be seen in the *Summary of findings* prototype UI and in the Google spreadsheet.

What is *perceived* as *effective* in terms of overall improvement?

This question dominated the process of extracting data to show "overall improvement." Although ACR20 is considered a "response" to a DMARD, it may not be sufficient for a patient to consider themselves significantly relieved. FDA guidance to manufacturers says that a claim of "major clinical response" requires statistically significant response of ACR70 continuously for 6 months in an adequately controlled clinical trial of at least 7 month duration (Food and Drug Administration, 1999). FDA cites data showing reference response rates at end of trial in a "comparative multicenter trial" of methotrexate (n=119) versus auranofin (n=118) of:

- **Methotrexate**
  ACR20: **65%**
  ACR50: **35%**
  ACR70: **9%**
- **Auranofin** [gold salts]
  ACR20: **29%**
  ACR50: **18%**
  ACR70: **6%**

Since even FDA does not consider a response below ACR70 to be sufficient to support a claim of "major clinical response," it seems that ACR20 is an unreasonably poor response to communicate to patients as "effective" in terms of overall improvement. Cochrane reviews of DMARDs routinely report estimates of effect for an outcome of ACR50 *at* certain point (e.g. 12 months after treatment begins), which is also below the FDA standard of major response. This outcome is rephrased as "major improvement" in the Cochrane review of methotrexate for RA; the estimate of effect from that review is 23% of patients showing ACR50 at 12 months (Lopez-Olivo et al., 2014). Corresponding evidence in reviews for other results—such as functional outcomes—provide context for what ACR50 response means for those more granular outcomes. Secondary analysis of clinical trials whose primary endpoints were ACR responses also provide context. For example, such an analysis of data from the FIN-RACo trial of multiple-therapy with conventional DMARDs and corticosteroids looked at work capacity and disability 5 years after treatment began, stratifying participants into groups by their response at 6 months. Patients who had ACR50 response at 6 months also had better work and

disability outcomes at 5 years than those who had achieved an ACR20 response or lower (Puolakka et al., 2005). Considering the prevalence of ACR50 response in systematic reviews, and additional context, it emerged as a reasonable endpoint to report for overall improvement. However, an unsolved problem is how to truly help patients make meaning of that figure.

There is considerable heterogeneity even among Cochrane systematic reviews in terms of estimates of effect and follow-up time for ACR50 response. Occasionally estimates are produced based on a long follow-up time—for example, 79% of patients seeing ACR50 at about 3 years for therapy with etanercept plus another DMARD (Lethaby et al., 2013). In other cases, the follow-up is much shorter—for example, 30% of patients seeing ACR50 at about 6 months for therapy with tocilizumab plus methotrexate (Singh et al., 2010). That presents immediate challenges for communicating clearly to patients about these medications: Does one have to *wait* for 3 years to see the full effect of a medication? Would the ACR50 response for tocilizumab plus methotrexate be sustained at 1 year, 2 years, or 3 years? Also, is 30% of patients seeing a 50% improvement a *good* outcome? Is 79%? In studies of how RA patients rated the utility of statements about medications in terms of "trading off" between alternative treatment options, the phrasing "75% (75 in 100) of people receiving this drug will feel much better" (mean score 61) received a much higher utility score than "45% (45 in 100) of people receiving this drug will feel much better" (mean score 7) (L. Fraenkel et al., 2004). The latter example is much closer to the typical ACR50 outcome data found for inclusion in the prototype, which might be a concern. People might find the "truth" about the medication's effectiveness—as measured by ACR50—to be *not useful*. PROs for improvement might be a preferable alternative, should they come to be used more frequently.

It is obvious that any serious spelunking in the RA literature by a motivated patient would require some guesswork about the meaning of these measures. First, the population studied in each of the sources of evidence may be different—and in fact, they usually do vary. Biologic DMARDs are usually a second or third-line treatment, and estimates of their effects are thus usually derived from research on their use by people who have not achieved an adequate response with methotrexate or other first-line treatments. Every person has their own unique understanding of their body and well-being, and consequently of their experience of RA, if they have it. Thus each patient has their own mental model of benefit and harm, although it is likely a flexible model. Without aid, to help someone form a specific—and hopefully somewhat realistic, in spite of the subjective nature of experience—mental model for each measure, there is likely to be limited consistency and usefulness in the interpretation of data. In other words, one person might think 50% improvement means complete success while another might think it is insufficient—indeed, there is no right or wrong answer. Such interpretation depends on where someone is in their life, in their disease process, their social support system, and other context. Qualititative research about the *human experience* of these outcomes might prove useful.

Presentation

Outcomes are evaluated at a certain point in time (follow-up time). These follow-up times may be heterogeneous from study to study, since there are not necessarily standards for the evaluation of all outcomes. While the follow-up

time may not be of interest to patients for *all* outcomes, for some they may be essential to a full understanding of the evidence. For example, if one wants to know both *when* a medication might take effect and *how long* a particular benefit (e.g. 50% improvement) might be sustained, reporting data about that outcome at various time points–say, 3 months, 6 months, 1 year, and 2 years–may be appropriate. In other cases, simply knowing that an outcome had occurred *by* a particular time is adequate–for instance, evidence that within 1 year, 5 people out of 100 in a clinical trial would have stopped a medication due to a serious side effect.

Overall improvement data from systematic reviews reported ACR50 at a variety of follow-up times, sometimes as a range. In a review of etanercept, the estimate of ACR50 was reported with trial follow-up times as a range from 24 to 156 weeks (Lethaby et al., 2013). When such a range was reported, the prototype uses the upper bound. For the purposes of this visualization, it is assumed that treatment effects are cumulative–for instance, *by* the upper bound of the follow-up range, a certain number of trial participants would have experienced the ACR50 outcome.

In order to illustrate this outcomes "over time," an abstract schematic *outcome timeline* (see Figure 3-5) was designed to plot outcomes on a horizontal display. In order that the icon arrays could remain large enough, the timeline is large–

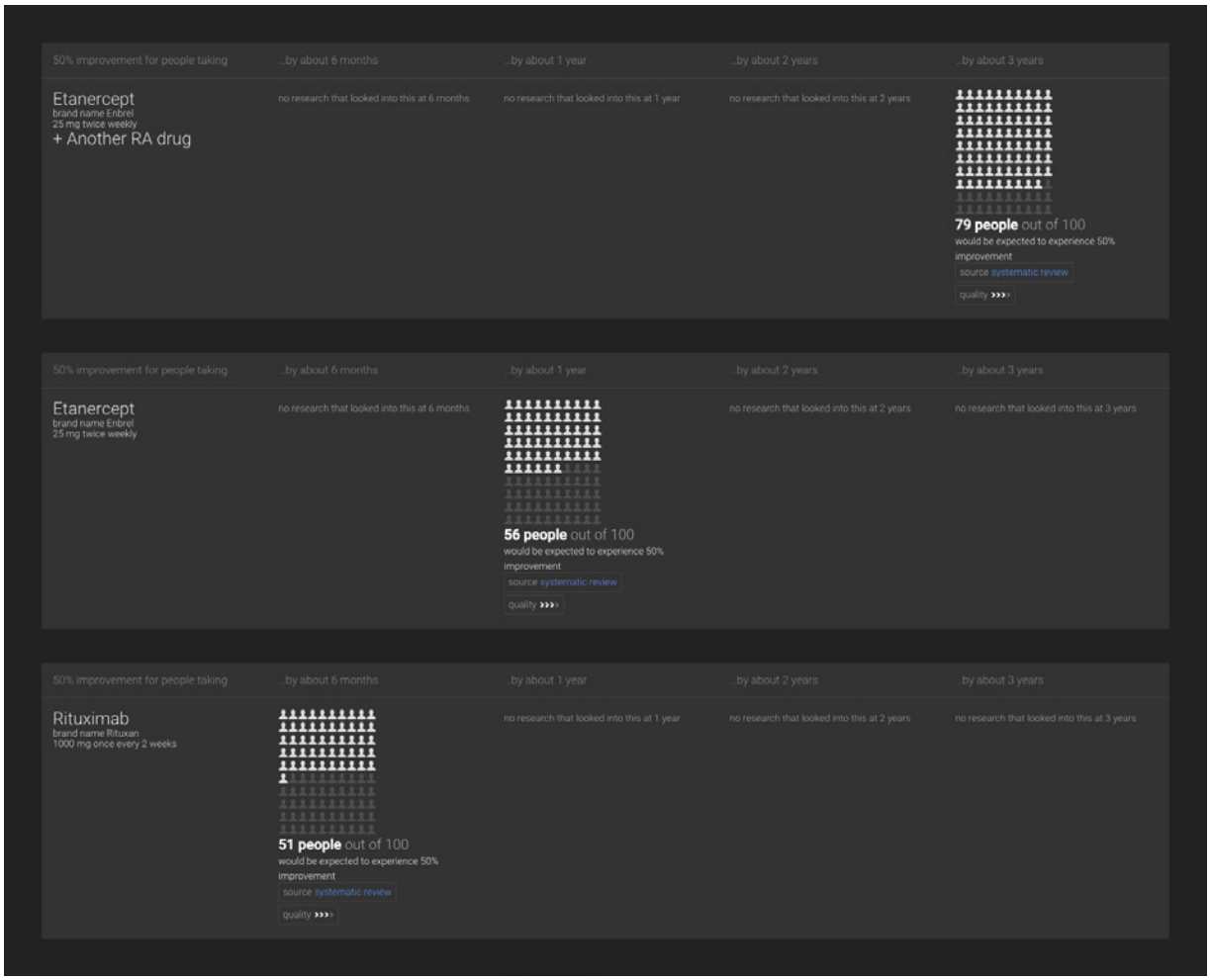


Figure 3-5
Overall improvement visualization in the *Navigator* prototype.

potentially indicating the need for a large display on which to visualize multiple medications. The *Navigator* prototype UI is responsive, meaning that it adapts to different display sizes. On smaller displays, the timeline is collapsed so that milestones (time points) at which there are no data for a given drug, are hidden.

Pain

Like fatigue, pain is a pervasive and cross-cutting symptom of RA. Even among patients who say their RA is "somewhat-to-completely controlled" a large proportion are dissatisfied with pain control, and those people report greater fatigue and poorer outlook (Taylor et al., 2010). Inadequate pain control is associated with worse psychosocial health status; it is also a factor that can be manipulated to have a positive effect on psychosocial health (Courvoisier et al., 2012). Pain is also a top treatment priority for RA patients (Sanderson et al., 2010). Treatment with anti-inflammatory drugs—such as NSAIDs like ibuprofen or celecoxib—may have an effect on symptomatic inflammation and pain. But treating the RA disease process itself with DMARDs should also have an effect on patient pain, so it is reasonable to report data on their effect on pain. After all, other pain relievers may be undesirable to or contraindicated for some patients. Pain is a by nature a patient-reported outcome (PRO) but rarely the *primary* endpoint of studies about RA treatments. It is part of the ACR core set, as mentioned in the prior section on overall improvement outcomes.

Working with evidence about pain outcomes

In many studies, including reports from trials included in Cochrane systematic reviews of DMARDs (For example in Lopez-Olivo et al., 2014), pain outcomes (usually recorded a 10 cm or 100 mm visual analog scale or VAS) are reported using a mean difference statistic: The difference between the mean pain score for people in an intervention group and the mean pain score in a control group, at a certain time point, such as 6 months after starting treatment . In some trials, mean *change* in pain score (not difference between intervention and control) is reported, sometimes with initial baseline scores (For example in Smolen et al., 1999). However, sometimes both these *and* other statistics are used, even in systematic reviews (for example in Steiman et al., 2013), making extraction of findings more difficult. Mean change or mean difference on a VAS may be common ways of reporting pain outcomes, but they are not the only ways: A review of 241 trials of interventions for fibromyalgia found 75 pain-related outcomes (Reported in Busse et al., 2015).

In an article on behalf of OMERACT recommending strategies for reporting pain outcomes in clinical trials, Busse et al. (2015) ask important questions related to the meaning of measuring and reporting on pain to patients, such as: "10 mm on a 100-mm visual analog scale (VAS) for pain may be statistically significant, but is it important to patients?" (Busse et al., 2015). They point out that because not everyone will experience that "average" effect, it may be more beneficial to communicate the proportion of patients who "report an important reduction in their pain." But then, what is *important?* That depends on the patient, obviously, and their experience of pain and how it interferes with their life. As Busse et al. note, 10 mm on a 100 mm scale is considered a minimally important difference (MID) based on studies of the instrument, but that may

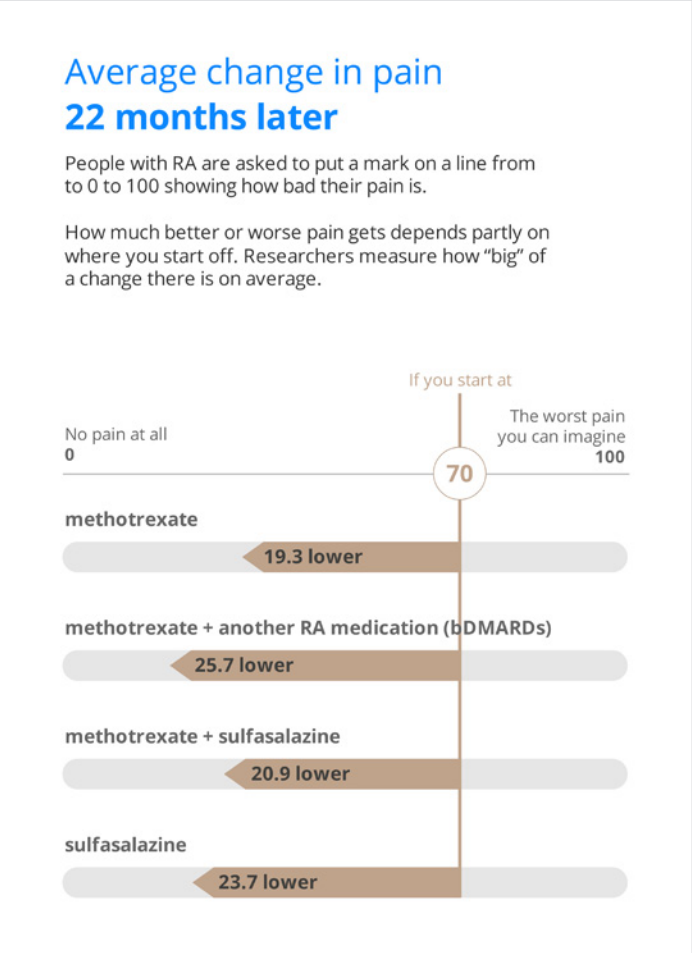


Figure 3-6
Sketch of interactive change in pain visualization. Data for this sketch from Steiman et al. (2013). The methotrexate estimate is the center of the range reported in those data (20 mm to 31.4 mm lower). The "if you start at" line is fictitious; the data do not report change from that specific baseline value.

not map to an important difference for patients: "In the absence of consensus on what constitutes a patient-important threshold in pain relief, it is reasonable to provide a range of options. To provide guidance in this regard, participants of the 2014 OMERACT Workshop advocated for reporting either an appreciable reduction from baseline pain (e.g., 20%, 30%, or 50%)," or a number of dichotomous outcomes, which in investigating evidence for the purposes of this prototype were rarely encountered (Busse et al., 2015).

The prevailing format of pain-related outcomes encountered during research for the prototype were mean changes on a VAS. Many of the challenges in harmonizing data reported using other scales (like effect size) with VAS scales, including conversion to a meaningful standardized unit, have been deferred to future work with the prototype. Two pain-related outcomes are supported–mean change (from baseline), and difference in mean change between placebo and intervention (*mean difference*). In the prototype, a simple process standardizes findings as an overall estimate of change in pain in MID units. All findings must be represented as a mean change. Findings that report a mean difference must have a reasonable placebo mean *added* to the difference, so that it becomes an overall mean change from baseline, no longer a difference. Instead of taking the placebo mean from the source estimate, all placebo means for all intervention vs. placebo comparisons are pooled, and their unweighted mean calculated–the *pooled placebo mean change*. (That assumes–perhaps naïvely–that placebo estimates are reasonable to compare from source to source.) The pooled placebo mean change is added to the intervention's mean difference value, turning it into a final mean change in score that can be converted to MID units. A more sophisticated method would consider sample sizes, variance, heterogeneity (e.g. of dose, in duration of follow-up, or population), and perhaps pool findings for each intervention to determine an overall effect size for that intervention–effectively, automated meta-analysis. For now, each individual estimate is simply reported as a mean change in pain as MID units.

Presentation

The need to show change in a standardized unit–which could be MID, standard deviations, or some other metric–led to the development of a generalized display loosely based on a visualization in certain decision aids published by the Mayo Clinic (See the Depression Medication Choice and Diabetes

Medication Choice decision aids from Mayo Clinic Shared Decision Making National Resource Center). Earlier sketches (see Figure 3-6) explored displaying change in pain on an interactive representation of the visual analog scale. On that design, the user might specify an initial (baseline) pain score, and see the relative change with each medication. However, that presented obvious challenges since the estimated change in pain might extend past the scale's edges if someone felt they had especially mild pain and the medication effect was large. Such an interactive scale would require data about the relative strength of effect given different baseline pain levels.

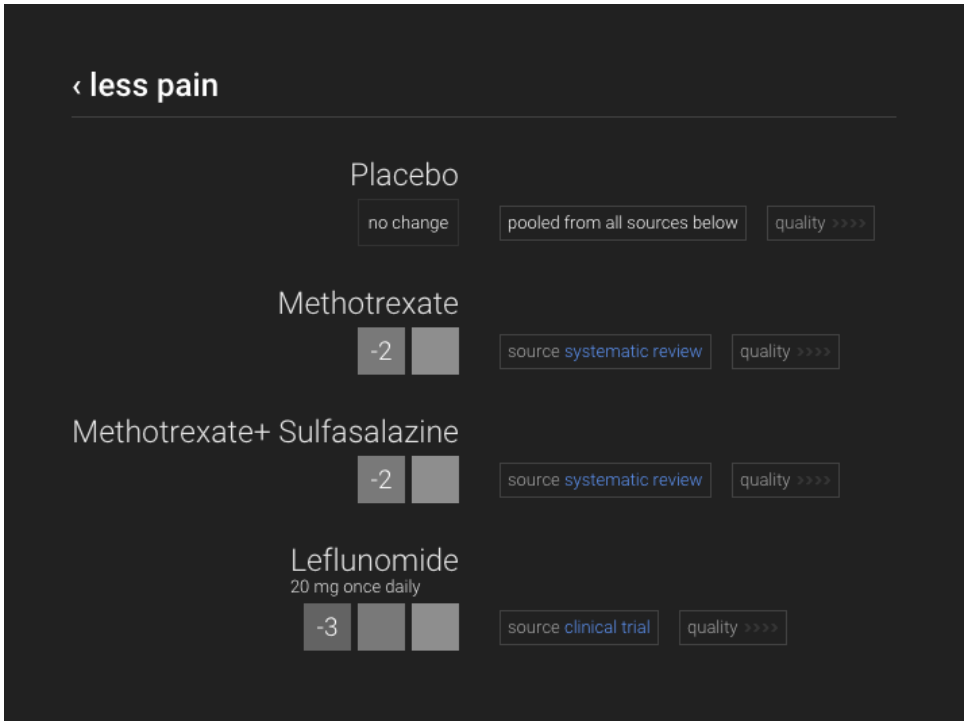


Figure 3-7
Change in pain visualization in the *Navigator* prototype.

Instead of an interactive display, the current visualization (see Figure 3-7) shows the expected change in square blocks representing MID units. The expected placebo change is at the top of a vertical list. Each intervention can be compared to the other interventions visually. Fine differences (less than 1 MID) are omitted visually to simplify the presentation. Although this visualization is here described in terms of change in pain scores, it can be adapted to show changes in any standardized unit, and the MID customized.

Work

Although "working" is not necessarily a top treatment priority for RA patients, enjoying life and being able to do everyday things are (Sanderson et al., 2010). For someone who has been forced to stop working, because of fatigue, pain, loss of function, or depression, returning to "normal" may include being able

to work. It also may mean the capacity to earn a living and be financially stable. When it comes to any medical treatment, a reasonable question is: Does this treatment help people live the way they want to live? This is absolutely the kind of outcome that is exceedingly difficult to measure. But if medications to treat RA are evaluated by their ability to reduce symptoms *and* slow disease progression, perhaps there is an association between primary outcome measures—like the ACR core set—and quality of life outcomes like ability to work. And perhaps communicating such evidence could give people an idea of what they might expect not just in the next few months, but the next few years.

Evidence about work-related outcomes

Predictably, there are myriad ways to measure work-related outcomes. They include employment status, absenteeism, presenteeism (attending work while ill), days worked, years worked, likelihood of stopping work, hours of productivity lost, utilization of disability resources, retirement rate, work productivity impairment, and other outcomes. These may be patient-reported (e.g. self-reported impairment) or gathered from other sources, like government data about disability benefits distribution. Perhaps because of the difficulty in teasing apart the effects of individual medications themselves on work-related outcomes, it is easier to find research where such outcomes were analyzed secondary to usual RA outcome measures, and where medications are grouped and compared—for example, looking at biologics generally versus conventional DMARDs.

Out of 12 systematic reviews in the Cochrane library which were used as sources for data for the DMARDs included in the prototype, only the one on methotrexate included analysis of a specific work-related outcome (Lopez-Olivo et al., 2014). Other studies seem more representative of work-related evidence. For example, secondary analysis of data from a trial of combination therapy in a cohort of Finnish RA patients (Puolakka et al., 2005) found that patients who saw the greatest benefit from treatment at 6 months—remission or ACR50—had lower work loss 5 years later, regardless of the therapy. Eriksson et al. (2013) compared change in work loss for people with RA who did not respond well to methotrexate, and were then randomized to combination therapy—biologic (infliximab plus methotrexate) or conventional (hydroxychloroquine, plus sulfasalazine, plus methotrexate). Treatment adjustments were permitted: "Both sulfasalazine and hydroxychloroquine could be discontinued and replaced by cyclosporin A (2.5 mg/kg/d in divided doses; increase allowed to 5 mg/kg/d), and infliximab could be discontinued and replaced by etanercept (50 mg/wk)" (Eriksson et al., 2013). Prior to randomization, the mean days of work loss per month for those non-responders was 17 (SD 13). This study found that *either* kind of combination therapy decreased the median days of work loss over time, with the maximum improvement seen between 8 and 12 months after randomization. By 21 months, the group receiving conventional treatment fared slightly better, but the authors concluded that there was no significant difference between treatment groups.

While the OMERACT collaboration has helped standardize many RA outcome measures, the heterogeneity in work outcomes and related study design makes working with such data more difficult. A systematic review of the effect of biologic DMARDs on work participation concluded that biologics probably show positive results on absenteeism and presenteeism compared

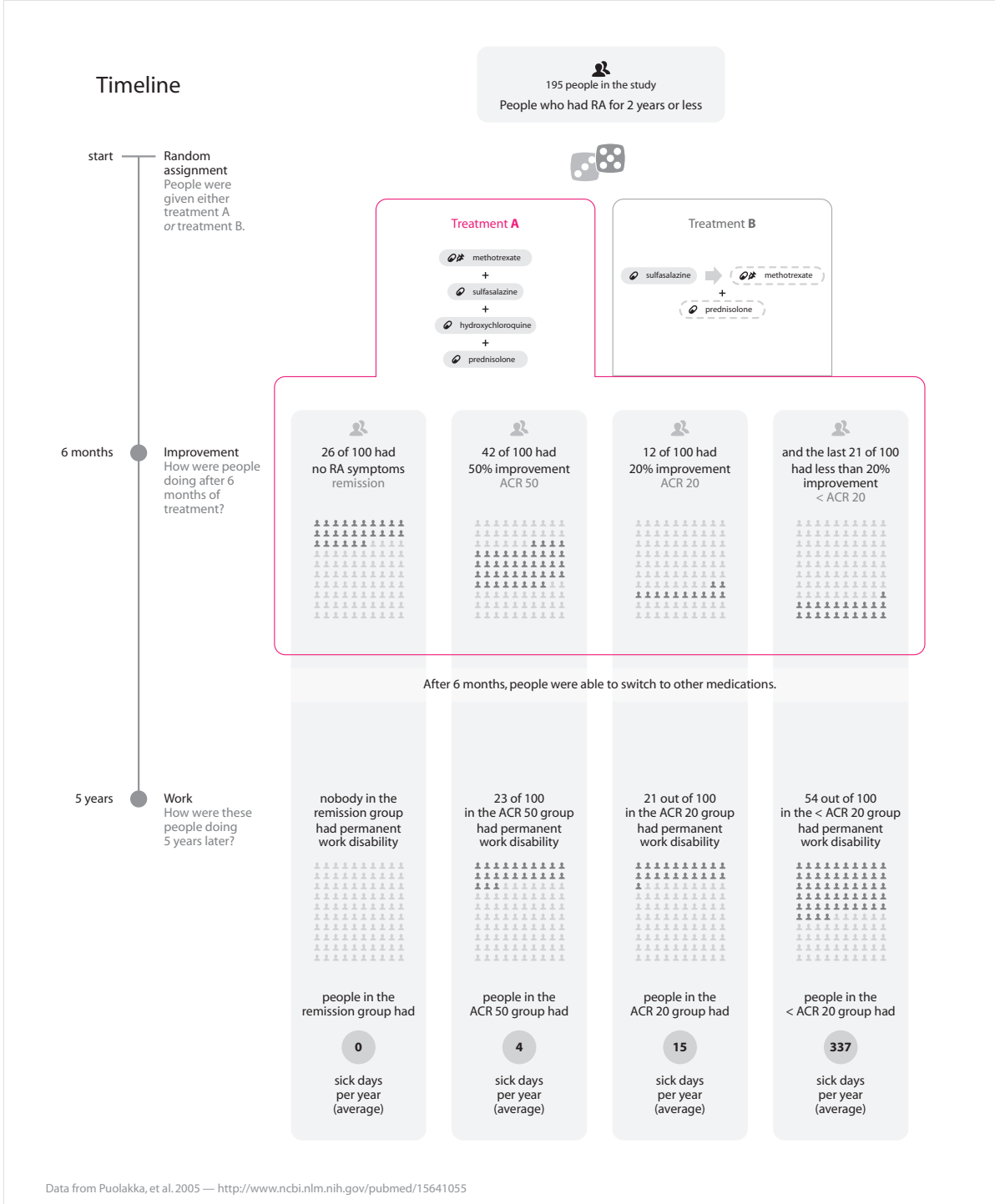


Figure 3-8
Sketch of a work-related outcome visualization. In this visualization, the medications themselves are deprioritized in favor of explicating that the work-related outcome at 5 years is secondary to a primary overall improvement outcome at 6 months. Data from Puolakka et al. (2005).

with usual care "although no pooled effect size could be calculated due to the heterogeneity of all data "(Wee et al., 2012). The authors discuss shortcomings in the included studies that hamper comparability, including that work participation was a secondary outcome in all but 2 of 19 studies, that populations varied significantly from study to study, and that the evidence left unresolved questions like whether biologic DMARDs have additional benefit compared with intensive DMARD treatment since "few of the studies in this review compared TNF with intensive DMARD treatment." They conclude their review by saying: "To enhance the comparability of studies, consensus on preferred outcome instruments and recommendations on the conduct and reporting of studies on work participation is recommended."

One additional wrinkle in looking at these outcomes is that work participation at diagnosis or start of treatment may be one of the best predictors of later status (Olofsson et al., 2014). It may be that effective communication to patients about how RA treatment might help them sustain, improve, or return to work has to do with tailoring evidence to their current status. Regardless of how well treatments work to improve symptoms, it may be that someone whose RA has driven them to leave work may have difficulty returning even after improving.

Implications for the prototype

The upshot of the evidence on RA treatments' effect on work seems to be that a good response to medications (ACR50 or better), ideally *before* RA has led one to reduced employment or productivity, is associated with decreased work disability over the longer term—regardless of the specific treatment. Attempting to find and work with evidence about work participation was challenging, providing a forceful counterpoint to the hypotheses of this project itself, and to methods chosen to work with other kinds of evidence. First, there may not be a sufficiently broad and clear base of evidence to draw from. Second, the evidence may not be easily encodable—due to heterogeneity, a much more sophisticated schema might be necessary to encode all the relevant details. Third, the data might be difficult to harmonize and display for similar reasons: It may be that the populations, study designs, and outcomes are simply too different to report in a coherent and consistent way, as per Wee et al. (2012). Fourth, individual medications or combination therapy may not matter much, although the prototype is entirely geared towards findings about individual medications. It may be that evidence generalizable to "biologics" could be "copied" to each medication, but that might be disingenuous in presentation,

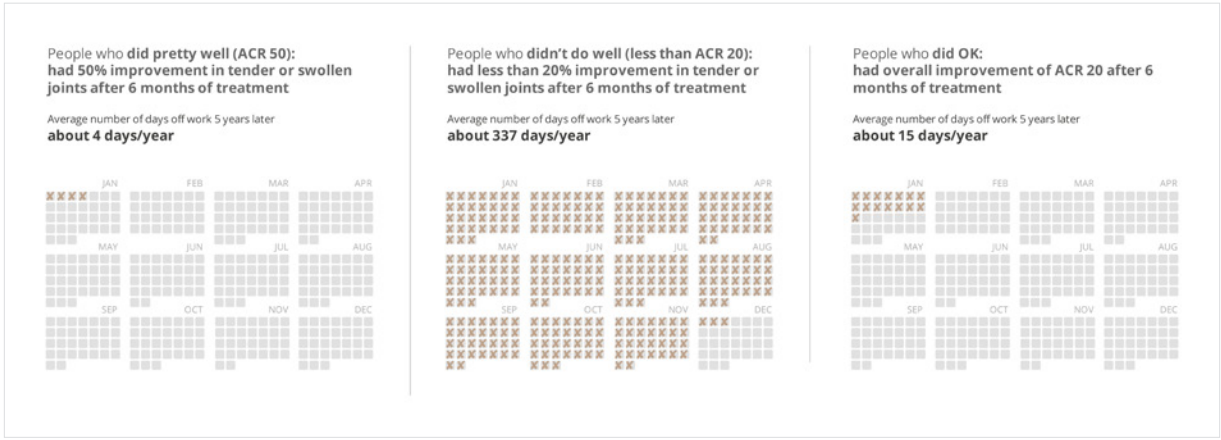


Figure 3-9
Sketch of an alternative work loss outcome visualization. Data from Puolakka et al. (2005).

as if there was research on each medication when in fact there was not. Ironically and to summarize the counterpoint to this project, it may be that a few sentences more clearly communicates the upshot of evidence on work participation.

At the time of writing, the prototype contained very little evidence about DMARDs' effects on work participation and status, and the question of how to combine heterogeneous data remained unresolved.

Presentation

In order to present evidence about work-related outcomes, the prototype probably needs to accommodate and clearly explain the "grouping" of treatments. In other words, whereas the other major *issues* are reported on a per-medication basis, somehow data would need to be grouped to show that it applies to *all* biologics or to conventional DMARDs, generally. It also probably needs to similarly accommodate and display outcomes secondary to a given primary outcome. In that case, users might be prompted to understand what they might expect if *whatever* treatment they choose leads to remission, vs. ACR 50, vs. ACR 20, etc. In that way, it may be that this kind of secondary outcome evidence would help answer the question, "what does it mean to have a 50% improvement?" Sketches for communicating evidence about work participation explored these possibilities, but the prototype has not yet been extended to support them (see Figure 3-8 and Figure 3-9).

Side effects

Intuitively, knowledge about potential harms of medications—side effects or adverse effects—are essential to informed medical decision making. Like any pharmaceutical products, DMARDs can produce side effects, bothersome and serious. Relatively non-serious side effects may play an outsized role in patient preferences for RA medications (L. Fraenkel et al., 2004). Adequate balanced presentation of evidence about side effects is desirable because of their importance to patients in medication choice. The side effect profile of a medication can give someone a sense of what to expect from it, perhaps more so than focusing on a single side effect. However, because of heterogeneity in both study and reporting of side effect data, these may be particularly difficult to harmonize and present in a non-misleading way. Consider the following principles which might underlie effective presentation of side effect data:

- Show that the side effect *is* a true side effect of the medication, rather than a common experience of humans in general, or people with RA (i.e. that it occurs at a rate statistically greater than placebo). Good data would disambiguate RA symptoms from medication effects, and from baseline human experience.
- Show how common the side effect is (i.e. incidence or frequency).
- Communicate likely onset and duration of the side effect (i.e. when it is most likely to occur and how long it tends to last). Some are associated with each dose, some with the body adjusting the medication, while others might happen after a long duration of use.

- Communicate severity or intensity of the side effect, including whether or not that intensity is dose-dependent. Put another way, how much the side effect interferes with feeling well or living well, and whether that effect varies with the dose.

With such information, one could learn that one medication tends to cause intense nausea in half of people who take it for the first few days, but that it tends to go away. Another might increase the risk of non-serious infection slightly but only with use for a year or so. Or another might cause rarely cause slight dizziness, but if it did, cause it with each weekly dose and never really go away. Such information is a kind of guidance about what to expect with the medication. Unsurprisingly, the great heterogeneity in side effect data means that answering such questions is rarely possible. A brief discussion of certain sources of side effect data illuminates this problem.

Sources of side effect data

Sources of side effect data include product monographs (Structured Product Labeling or SPL), registries of patient data, post-marketing surveillance databases (including electronically through openFDA), patient-reported outcome data (Web sites like PatientsLikeMe for example), post-marketing clinical trials and systematic reviews or meta-analyses of such post-marketing data. Each source has weaknesses in terms of biases related to data collection and reporting, intended audience, and access.

Product literature. An obvious starting point for side effect data is literature supplied by pharmaceutical companies. Such literature summarizes side effect (adverse reaction) data from clinical trials of drug products before they are brought to market, and updated with some data about side effects observed *after* they are brought to market (post-marketing data). There is considerable variation in the ways that side effect data are studied and reported in product literature, partly owing to flexibility afforded manufacturers by regulators (See Guidance for Industry on Adverse Reactions in Food and Drug Administration, 2006). These sources report adverse effects heterogeneously enough so as to make aggregation both impractical and potentially misleading. Data may be trapped parenthetically in sentences, non-systematically reported, with or without incidence rates. Even when data are presented in a table, they are not semantically encoded and easily machine readable. Even more concerning are data disparities which do not make comparison from monograph to monograph straightforward.

For example, the SPL for etanercept ("ENBREL labeling," 2015) reported adverse effects in RA patients observed during the course of relatively long clinical trials—from six months to two years. The SPL for tocilizumab ("ACTEMRA labeling," 2014) reported adverse effects in RA patients from trials of 24 weeks' duration—even though the labeling itself indicates that thousands of patients were studied for much longer, up to 3 years. In the etanercept SPL, infection rates (bacterial, viral, or fungal) in the 2-year study were 81% for people taking etanercept vs. 86% for people taking methotrexate (the comparator), and 50% for etanercept vs. 39% for placebo in 6-month studies. In the tocilizumab SPL, infection rates were not pooled together, and reported for the 24 week trial duration. For tocilizumab monotherapy at a dose of 8 mg/kg, upper respiratory tract infections (7%), nasopharyngitis (7%, common cold),

and bronchitis (3%) were reported separately and do not differ significantly from reported rates among patients treated with methotrexate alone or placebo infusion plus another DMARD. Nasopharyngitis and bronchitis are plausibly outcomes resulting from viral or bacterial infection, but not indicated as such. If none of those three outcomes occurred concurrently for trial participants, the true rate of any infection could be 17% or higher. But because the side effects were reported individually, there is no way of knowing. The two SPLs present side effect data that are difficult to reconcile.

An ordinary person might consider respiratory infections, the common cold, and bronchitis to be similar—although perhaps a common cold to be less serious. The same might go for muscle pain, muscle weakness, musculoskeletal pain, and joint pain, which may be difficult to disambiguate. Nevertheless, the pharmaceutical company can effectively obfuscate the true incidence of side effects by reporting these separately and on their terms instead of in patient terms. There is an inherent risk of bias or distortion in that the research is usually funded by the manufacturer, and submitted to regulators as part of the drug approval process, to prove that the drug is safe. Moreover, terms like *nasopharyngitis* and *asthenia* are technical. A controlled vocabulary to translate or group these terms into patient-relevant language would be necessary, or they need to be manually translated at the time of encoding into a data source like the spreadsheet for the prototype.

Registries. So-called "patient registries" are databases of observations of patients in the real world. They vary in design: Some may feature representative samples of patients with a single condition and detailed longitudinal data, while others may have limited data for much more diverse populations. Particularly for adverse effects, they may be a more reliable source of data than product literature and clinical trials, since they often allow researchers to "follow" patients for a greater length of time (years instead of weeks or months), and may feature more consistent reporting of side effects from drug to drug. A number of RA patient registries exist, with surprising variation in patient comorbidities, average RA disease activity, and other characteristics (Curtis et al., 2010). While registries might be a good source of side effect incidence, their data are not generally openly available. Because research agreements might have been required to access data, they were not pursued as data sources for the prototype.

Post-marketing surveillance databases. National or supra-national regulators like the Food and Drug Administration (FDA), Health Canada, and the European Medicines Agency, operate surveillance systems to monitor the safety of medications once they have been approved and are for sale. Access and machine-readability of such data varies. Of these three agencies, only FDA supplies machine-readable data suitable for integration with software like the prototype (Kass-Hout et al., 2015). Although openFDA provides an application programming interface (API) for real-time inquiry of up-to-date adverse reaction data, the format of the data limits their value in the prototype. For one thing, it is not possible to query a priori for results in which the medication in question was the primary or suspect drug; such processing must be done on the whole set of reported adverse events. Furthermore, data are unvalidated, causal relationships between the drug and reaction need not be proven for a report, reporting is voluntary, and there is no way to estimate the incidence of such side effects in the population at large (See Drugs API reference on openFDA, FDA, n.d.).

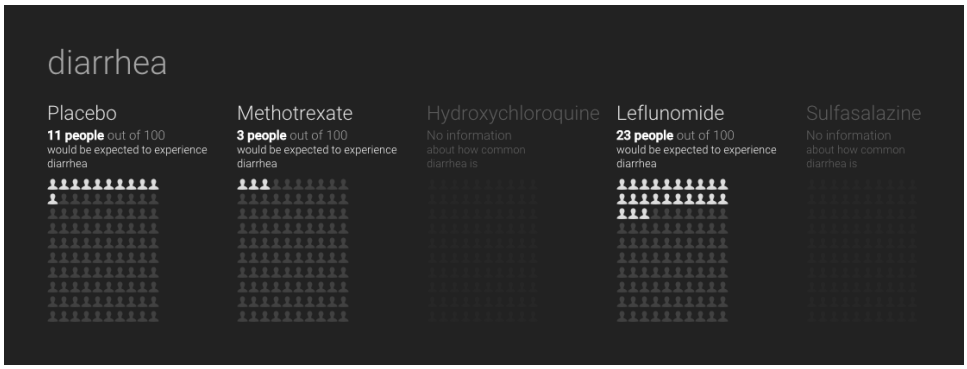


Figure 3-10
A side effect visualization in the *Navigator* prototype.

Patient-reported data online. Web sites like PatientsLikeMe have patient-reported side effect data, though they are not made available to external developers (see *Online consumer health resources* in Chapter 2). These data benefit from being reported on patient terms and in a naturalistic setting (the real world), but they are unvalidated. As such, they might well complement controlled study data.

Post-marketing clinical trials, systematic reviews, and meta-analyses. Research specifically investigating the safety or adverse effects associated with RA medications might establish individual side effects as endpoints of interest. From a clinical perspective, though, it may be more important to investigate serious or life-threatening side effects than less-serious ones. Systematic reviews routinely report a broad serious adverse event outcome, and frequently on withdrawals from clinical trials due to adverse events, but not always on detailed but patient-important side effects. Data on individual side effects may be reported using heterogeneous measures, and with varying levels of detail on side effects, not unlike product literature. A thorough 291-page review of Leflunomide for RA (Osiri et al., 2003) devotes space to a few specific side effects or groups of side effects: Alopecia (hair loss), elevated liver function (sign of liver injury), GI symptoms generally, allergy or rash, hypertension, weight loss, and infections generally. A more recent review of methotrexate for RA (Lopez-Olivo et al., 2014) has data for 28 individual adverse reactions, ranging from diarrhea to stroke and even death.

Working with side effect data

In order to compare the relative likelihood of side effects with one medication or another, sources that provided an estimate of incidence were preferred. The outcome measure is dichotomous—the side effect occurred, or has not occurred, by the follow-up time. Onset, intensity, and duration of side effects were not reported features of side effect data. Like other data, side effects were manually encoded into the spreadsheet. Purely because side effect naming (and therefore identification) is so heterogeneous, these data call out for at least limited harmonization. Variation in follow-up time, population and setting, and dose are also sources that indicate the need for harmonization. The primary metric—frequency of side effect—is generally consistent, so converting to other scales or measures is generally unnecessary. Occasionally they are reported in relative terms, as a relative risk or odds ratio versus a competitor.

Grouping side effects into patient-important clusters (for example, those suggested in L. Fraenkel et al., 2004) requires both an ontology and a method for determining an appropriate frequency to report. Without a comprehensive

ontology, aligning side effect terms (e.g. gastrointestinal distress with GI symptoms with stomach upset) is a non-trivial challenge. It is important to know when terms are synonymous, and when terms subsume others—for instance, that *infection of any kind* encompasses *upper respiratory tract infection*. Even if side effects can be aligned or grouped it is not clear which frequency of which side effect is an appropriate estimate. The highest reported frequency might be a rough guide—so as to say, "up to 30% of people taking this medication got an infection of any kind." A range may be more appropriate, if harder to interpret. Displaying frequencies for a larger number of more specific side effects may mislead by appearing to underestimate the chance of experiencing a side effect, but more accurately represent the real likelihood.

Since the prototype does not employ an ontology, the chief harmonization challenge is one of follow-up time. Given the etanercept vs. tocilizumab example above, it is misleading to report side effects without a notion of the follow-up time. However, a good solution to this problem remained unresolved at the time of writing, since follow-up time in data about adverse events was so heterogeneous. Ultimately, a mixture of sources—including product literature and systematic reviews—was used to populate the prototype.

Presentation

Since data to show onset, intensity, and duration of side effects had not been found for the prototype, only reported frequency is displayed using an icon array visualization as with data on overall improvement (like ACR 50). One side effect can be viewed at a time in the *Navigator* prototype (see Figure 3-10 for an example). Each medication has its own "slot" to display data for the selected side effect. In the event there is no estimate of the frequency of that side effect for a given medication, that is clearly indicated and the icon array is dimmed. This is an area of the prototype that would benefit from devoted improvement.

Technical architecture

There is a proliferation of contemporary and free-to-use software development technologies and tools. They range in scope from programming languages, to databases, to instrumentation systems, to frameworks and libraries that speed development, permit communication between software systems, or facilitate creation of certain parts of software. Such libraries may be used for converting data from one format to another, or controlling user interface elements, or creating data visualizations. These may be combined in virtually limitless ways, as needed, into what is called a *stack* of technologies—a software project's architecture from soup to nuts. For any software project—and particularly those whose source code is to be shared with a wider community—the choice of tools is important, because it to some extent dictates and constrains the potential

of the software. Projects written esoteric programming languages will be less maintainable and readable by other developers. Projects that employ a less popular platform may benefit from fewer helpful and ready-made libraries, which means that their developers must shoulder more of the programming burden—sometimes reinventing the wheel, as it were. While a thorough discussion of technology-choice tradeoffs is beyond the scope of this thesis, it is an important feature of the prototype, so a brief elaboration of the specific choices follows.

Programming language and platform

The prototype is written primarily in JavaScript, a popular programming language used especially widely for Web-based software. JavaScript is a good choice because many parts of the prototype, from its underlying platform to user-visible data visualizations, can be written with it, and benefit from ready-made frameworks also written in JavaScript. Thus fluency gained while working on one part of the prototype was easily transferable to work on other parts. JavaScript also smooths interactions between data and logic; in most cases data *not* in the spreadsheet in the prototype is represented in JavaScript Object Notation (JSON), a format readily processed by logic written in JavaScript, and also widely used on the Internet for transfer of data between services. JSON is a "lightweight" key-value pair data representation format, simpler than alternative data formats like XML—indeed, it has been called "the fat-free alternative to XML" (Crockford, 2006). Basic medication data (from JSON) and spreadsheet data (outcome data) are processed and converted to JavaScript objects by the prototype.

The prototype is built with Node.js (Node) as the underlying platform or run-time environment. Node was chosen because it is open source, its primary use case is Web-based applications, Node applications are written in JavaScript, and there exists a significant repository of Node-based open-source libraries and frameworks to speed application development. For example, the prototype uses Express.js—a Web server—to "listen" to HTTP requests and "serve" or respond with Web pages from the prototype. It also uses lodash, a library to speed up data manipulation.

Data encoding, storage, and retrieval

One of the most significant questions early in the development process was that of what technology to use to encode data, and how it could be retrieved by the prototype. A few dominant concerns governed the choice of Google Spreadsheets over other options—mainly a database technology like MySQL, PostgreSQL, or MongoDB, storing the data in files (such as text files), or encoding or storing the data directly in the software, alongside logic. These concerns revolved around ease of development, and satisfying one of the prototype's key objectives: To demonstrate ease of adding, updating, and ultimately using data. To choose a convenient or reliable technology, but one which did not satisfy that important requirement, would be inappropriate.

Briefly, the most important requirements for data encoding and storage included the following:

- Data should be quickly and fluidly entered, ideally using a familiar user interface.
- Some structure, to guide input, could be easily imposed, but also changed easily.
- No programming skill should be necessary to add data.
- Multiple contributors should be able to collaborate and add data.
- It should be possible to annotate data.
- Data should be easily updated or added, with real-time reflection of the new.
- There is no need to build a user interface *de novo* to manage data input and editing.
- One system should accommodate all the data necessary for the prototype.
- The system should be free of charge.
- Data should be easily exported, should needs change.

Several options and dimensions were considered to arrive at a minimally effective solution.

Data in code. Situating data in the software code itself, alongside data, is generally unsuitable because it would be difficult for non-specialists to use. Although it would be checked in to a source-management system (GitHub), it would generally require some programming skill, and would not satisfy the requirement for fluid data entry using a familiar user interface. In general, data would be stored in text files and highly structured—perhaps in JSON format, which features an idiosyncratic and strict syntax—but not in a way that is easily changed or annotated. In such cases, it would be nearly impossible to "see" or filter an overview of data. Depending on the organizational scheme chosen, it could also be that the data themselves could become entangled with or encroach on the logic code, which would make it difficult to disentangle and share reusable parts of the code *without* the data. Basic medication data *is* stored "in code," though in its own file, separated from logic.

Text files. Storing data in text files suffers from most of the same problems as situating it in source code itself; in fact, it is a nearly identical solution except that there would be a strict separation of logic (code) and data. There would remain the same problems of accessibility, organization, format, and so on.

Databases. At first glance, it would seem that many database technologies satisfy the requirements outlined above. However, there are more arguments against them, chief among which is that there only rarely are accessible user interfaces for editing data in databases. In most cases, graphical user interfaces for editing databases are designed for and used by administrators of the database; those interfaces serve as control or maintenance mechanisms, rather than authoring tools. For a database that would be hosted alongside the prototype, an additional tool would need to be found and hosted elsewhere, and special credentials provided to contributors—more work than is ideal. For commercial database services hosted by a third party provider, such as Mongo,

authoring access would depend on their tools, which are more complex than a spreadsheet, or a custom interface would have to be created from whole cloth specifically to support the needs of collaboration, annotation (generally not supported by any database administration user interface), and a flexible structure.

Schema. Indeed, the second major argument against using a database has to do with structure, or schema. Most database technologies require that the fields and overall structure and connections between data be defined a priori. That is not a universal requirement, but it is most common, especially among freely usable database technologies. The prototype required flexibility of structure; the "right" ways of organizing data needed to emerge as the prototype was developed. Thus a more restrictive, schema-requiring database technology was not ideal.

Cost, licensing, and other issues. The remaining significant issues with using a database technology include cost and licensing, ease of configuration and maintenance, quantity of data and protocols for retrieving and using it in the prototype, and capacity to export data. Since the prototype is intended as a demonstration of repeatable methodologies for creating useful interfaces with medical evidence, it was preferable to avoid any third-party services with notable cost or licensing restrictions, or (admittedly subjectively evaluated) difficult setup or maintenance schemes. Relatively "open" (readily licensed or integrated) and unrestrictive database technologies, like MySQL, have more onerous and complicated configuration and maintenance, while those featuring easier setup (for instance, services offering database hosting) can cost money (such as the provider Compose.io).

Quantity of data and protocols. The prototype uses a minuscule quantity of data—several dozens of kilobytes—that easily fits in working memory. The prototype's source data are also heterogeneous and richly interconnected by their relationship to a small number of entities of interest (RA treatments), with any given user activity requiring records throughout the dataset. In other words, a typical scenario benefits from the software being able to "work with" the whole dataset in memory, alongside the running prototype. If the prototype used a typical database model, it is likely that multiple queries would have to be made to the database in sequence, slowing the software down to the point that it could interfere with the user's perception of its speed; with the data in memory, it can be manipulated much faster, with minimal latency. Although a database technology could be used, and queried by the prototype to retrieve all documents to store them in memory, its other strengths would have to outweigh relative weaknesses, since the most obvious benefits (efficient storage and querying of large datasets) are practically insignificant for the prototype.

Data portability. Although it is preferable for the prototype's data storage technology to remain in service and useful for as long as possible, that it may not is a risk. The best way to mitigate that risk is to choose a technology that provides for data portability: A straightforward way to get the data out of the store, in a format that can easily be repurposed and reused. Some database technologies support export, using administrative tools, but in many cases software would have be written from whole cloth to do just the export. As in the immediately prior case, relative weakness in terms of data portability would have be outweighed by other major benefits to argue in favor of using a database technology; the most readily used database technology, MySQL,

supports this export by way of multiple queries and third-party administrative tools, but imposes some burden.

The solution: Data in Google Spreadsheets

During initial stages of the prototype's development, Google Spreadsheets (GS) was used for encoding and exploration of structures for data. It offers a familiar spreadsheets user interface, is free to use, and enables rapid entry of data, and relatively quick structural changes. When it came time to find a way to export or use data in the prototype software itself, it became clear that it was actually a suitable data store for the prototype, satisfying most of the requirements: The key discovery was that Google provides a JSON API that can be used to query a spreadsheet in real time, returning the data in its cells in a way similar to how a database might be queried. The working prototype thus uses GS as its data back-end, in lieu of a typical database technology. An example of a similar product using GS as a back-end—Snake Oil Supplements—was discovered during development and is reviewed in Chapter 2. The benefits of using GS are:

- An easy to use user interface that future contributors could easily use to add data to the prototype.
- Similarly, a user interface that supports collaboration and annotation and public viewing of the evidence.
- Fast data entry and real-time updates. Data or parts of the controlled vocabulary can be edited in the spreadsheet, with changes reflected immediately in the prototype.
- A readymade API for querying data and returning it as JSON, easily processed by the prototype.
- Data can be easily exported.
- Schema-free design, so that the data schema can emerge and change, which is necessary for a rapid iterative development process.

The very same reasons GS was an attractive tool for early data-encoding experiments make it a suitable choice for the prototype, especially as it is intended to demonstrate a workflow or model for non-programmer researchers, designers, and clinicians to collaborate. It features a user interface accessible in almost any contemporary Web browser and updated frequently by Google; the user interface exploits a common mental model (the spreadsheet) familiar to many computer users. Thus GS facilitates rapid data entry and editing. Moreover, unlike administrative interfaces for databases, GS supports revision control, real-time collaboration by multiple parties, and annotation of individual data elements—annotations that are visible in the spreadsheet, but do not have to be exposed in the software. The collaborative nature of GS extends beyond editing the spreadsheet: It also features a graphical form designer that creates Web-based submission flows that could be used to solicit data.

Multiple kinds of data are stored in the spreadsheet: Adverse event data, measure names and descriptions, metric names and descriptions, a controlled vocabulary of patient-centered issues, and more. The spreadsheet model supplies adequate structure. While it is possible for anyone working with the source data to make errors in data entry (about as easily as with any of the data storage options), and perhaps more likely to make schema errors (putting the

wrong kind of data in the wrong place or in the wrong format) they are easily corrected—though by whom is a question of collaboration and quality control outside the scope of this research paper.

Most importantly, because the total quantity of data used by the prototype is so small, it's suitable to store it all in memory; a quick workflow or pipeline from data entry or edit through to use in the prototype is both desirable and possible. The GS API enables such a pipeline: The prototype uses the API to query for all the data in all the spreadsheet, reprojects it into a useful JSON format, and stores it in memory. The query happens once—when the prototype is opened by a user—and returns all necessary data to the client (the user's browser). Subsequent filtering and manipulation is handled by the prototype logic, as needed. However, this model facilitates such rapid iteration that a value can be changed in a spreadsheet and immediately reflected in the prototype.

Because GS is a service provided by a third party, Google, there are a pair of notable risks to using it. First is that although the service is reliable, there is a chance that it (or its API) could suffer a service interruption or be changed, thus interfering with the prototype. In the case of an interruption, the prototype would temporarily hobbled, unable to retrieve data. A caching strategy could mitigate this risk—but was not developed due to time constraints—to ensure a recent "copy" of the data were always available to the prototype. In the case of an API change, the part of the prototype that queries for data from the spreadsheets might have to be rewritten to conform to the new protocol. The second major risk is that Google, who operates GS, could permanently disable the service (or begin selling it at a cost). In these cases, a new home would need to be found—or the cost paid. Luckily, because GS offers excellent data portability, it is easy to maintain backups of the full dataset in formats that can be readily moved into a new data store.

Front-end and user interface technologies

Contemporary Web-based applications designed for use on desktop computers and mobile devices tend to be built using (Hypertext Markup Language) HTML, CSS (Cascading Stylesheets), and JavaScript. HTML is a markup language that is used to describe the content and—along with CSS—appearance of Web pages. Although a typical Web page might be understood as a digital, interactive extension of the model of a printed page, a Web-based application as complex as Facebook (with all its attendant features, including streaming video, real-time chat, etc.) is ultimately no different than any other Web page. It is simply a more complex one, still reliant on HTML and CSS to describe its content and display. However, much of such applications' complexity comes from additional code, usually written in JavaScript. There is no single standard application framework for such Web-based software, although almost every Web browser renders HTML and CSS, and executes JavaScript; thus they are reliable foundational technologies.

The prototype does not demonstrate any novel or complex data processing or computation, nor is it intended to be used in any way other than as an example, so the most ordinary JavaScript is sufficient for its core logic. However, its user interface is the key part of the project, so greater attention was paid to finding

an appropriate front-end and user interface framework. A discussion of the details of these libraries and their strengths and weaknesses is far beyond the scope of this paper, but a brief summary of the chosen framework is germane to the orientation of the thesis as a demonstration.

Many user interface frameworks quite strictly separate the JavaScript code (which often performs logic to determine what and how to present in the user interface) from the HTML code itself. It is often the case that the logic produces output data that is "merged" into the HTML using a template system, in much the same way that a database of addresses might be used to do a "mail merge" and produce neatly designed envelopes. By way of analogy, the envelope design would be crafted in HTML and CSS, while some JavaScript code would collect the addresses from a database, correctly title the proper nouns and check the postal codes, and the template system would finally be used to take the output from that code and inject it into the HTML envelope design. While a common pattern, it presents several disadvantages for the prototype; the most significant is that it would be nearly impossible to neatly present a single file that demonstrates a whole user interface concept, such as an icon array visualization to show absolute risk. Instead, it would be more likely that two or even three separate files would be necessary. If an open-source viewer or potential re-user were to want to learn how the visualization worked or even to re-use it, they would have to piece together its function from three files.

A second feature of many Web user interface frameworks is that they require manual manipulation of the Document Object Model (DOM). Without getting too specific, the DOM is structured, hierarchical representation of all of the elements on a Web page. Manual manipulation effectively means that any part of a Web page that interacts with, or is touched by the JavaScript logic of the application, must be manually added, tracked, altered, or removed, as appropriate. For a complex interactive application, where user interactions might change hundreds or thousands of elements on a page, this can be onerous—even when some of the manipulation is automated. One of the most widespread such libraries, jQuery, depends on—and effectively is a technology for—such DOM manipulation. It is possible to write user interfaces using jQuery that minimize the code that is strictly about DOM manipulation, but it is still necessary. Such code is overhead, or a kind of "tax" that is not essential to the demonstration of the user interface; thus it is preferable to avoid it.

The solution: React.js for front-end programming

React.js (Facebook, n.d.) is an open-source JavaScript library for building Web-based user interfaces that effectively abstracts these problems, and by doing so, is a suitable choice for the prototype: It facilitates the creation of compact, self-contained components that stand on their own to demonstrate necessary logic and presentation. For example, the same absolute risk icon array visualization can be represented using React.js as a single file, neatly presenting what it does and how it works to an open-source developer. In a nutshell, React.js components are written in JavaScript, and produce (for Web-based applications) HTML as their output. To reuse the envelope analogy, the React.js component is both the envelope and the template engine: It is simply given the addresses, and all the necessary logic for lightly processing the addresses and outputting appropriately designed envelopes is in one, self-contained package.

Even more usefully for complex Web-based applications, React.js automatically manipulates the DOM. There is no need for programmers to manually manage elements on the Web page; the React.js code simply needs to describe what any element should look like and how it should behave, and all of the manipulation necessary to respond to user interaction is handled by React.js itself.

Chapter 4

Discussion

This project began with a naive optimism and wide scope. The iterative design and development process, which always included reviewing and working with sample evidence, revealed layer upon layer of complexity that called even the premise of the project into question. Counter-arguments abound. Some patients do not want or would not directly benefit from greater access to medical evidence. Data on patient-important outcomes can be difficult to source. Evidence is ordinarily not machine readable. Innumerable social and technical barriers hold back the provision of such evidence. Heterogeneity in reported outcome measures, study methodologies, sources, data quality, and even outcomes themselves—conflicting data about the effects of medications—is common. Data may not conflict *per se,* but may be uncertain—because of the balance of benefits and harms, or merely statistically uncertain. The challenge of harmonizing heterogeneous outcome data alone is enormous, let alone synthesis that may ultimately be required to truly make sense of them. Even straightforward data about fairly consistently measured outcomes (e.g. ACR50) must be translated into language or images that the general public can understand, itself a complete domain of inquiry. They may need to be framed in terms that a patient cares about, such as effects on daily routine, exercise, feeling well—*issues.* There is the also, of course, question of when and how a resource that solved some or all of these problems and presented medical evidence to the interested person would actually be discovered and used. In other words, the role it might ultimately play in medical decision making, patient education or engagement, and so on. There are countless ways in which the data might be misinterpreted or misunderstood. Rigorous evaluation of design, usability, and evidence-understandability is thus necessary. Together, these make up the tidal forces *against* data-driven resources to enable wider access to medical evidence—frankly, for both patients *and* clinicians.

While all of these concerns were tacitly acknowledged and at least superficially understood at the outset of the project, their significance was under-appreciated. It took a great deal of in-depth experimentation and reading to peel back enough layers of the onion, so to speak, for the magnitude of their force to become clearer. The path from patient interest to evidence and back to the patient again—identifying information needs, finding evidence, encoding it, manipulating it, designing presentation of it, and creating a user interface to retrieve it, is difficult.

One initial purpose of the project was to find a way to bring together evidence on *multiple* outcomes (e.g. pain, overall improvement, side effects, and patient anecdote) to "explain" how they might interact and bear on a practical question like *will this medication help me ride my bike again?* It is crystal clear that such a purpose was deferred in favor of merely trying to present a minimal set of outcomes (e.g. pain, overall improvement, etc.) in isolation—such that someone might try to integrate the individual findings themselves and figure out on their own how those outcomes interact and bear on the question of playing tennis. (A doubtful prospect, indeed.) The original purpose might have been a better focus for the user interface design challenges, and perhaps would have produced a more meaningful outcome. However, the line of inquiry that was pursued, and that led to the prototype that was actually developed, at least serves the function of clarifying just how tricky it is to build an application that works with medical evidence. Criticism of existing online resources—data-impoverished text-based Web sites like WebMD—might seem righteous and simplistic in light of the realities of developing a data-driven alternative. In fairness, there are good reasons for WebMD to be like WebMD. In this thesis, expressed skepticism of the status quo is borne of awareness of the potential for something better, not utter ignorance. Even if that something better is much richer and more sophisticated than the prototype as it stands today—which it would have to be. Thankfully, it is ready to support such future work.

Limitations

This thesis is peppered with descriptions of the project's limitations. It is worth mentioning more of them in detail. First and foremost, the prototype itself is not suited for use by patients with RA or their caregivers yet. It is accessible, because it is open source, but it is restricted from being indexed by search engines and therefore unlikely to be accidentally discovered by a casual searcher. It is ready for future evaluation by such an audience *as a prototype*, in a suitable research context, and by others who work in the production and design of health communication resources.

Independent work and limited skillset

As a self-directed interdisciplinary project, building the prototype required "hard skills" in several domains, none of which were fully exercised in service of the concept. (Such is the nature of working independently.) The prototype is severely bound by the limits of its author's attention and skill in these domains—including user interface design, programming, data manipulation, interpretation of medical evidence, data visualization, basic statistical literacy, and so forth. The challenge of communicating information about DMARDs' effects on pain—dealing with heterogeneous outcome data, even just from a similar suite of systematic reviews, presenting them in a statistically faithful

and comprehensible way—illustrates how these skills were stretched, in light of balancing work on the written portion of the thesis and other parts of the prototype. More time would have allowed deeper investigation and skill development.

No "option of doing nothing"

In all medical decisions, one option is doing nothing. In effect, that is a potential "medication choice" that is not represented in the *Navigator* prototype. In contrast, many decision aids explicitly include such an option to contrast potential outcomes from interventions. Such data could be useful as a comparator in the prototype platform. Or, pooled placebo information could be used to provide similar context—it is already available.

Unvalidated data

As only one person read, entered, and reviewed data, there is the possibility that data have been erroneously encoded into the Google spreadsheet. Figures could be wrong or mis-typed, incorrect metrics might have been specified by accident, or the wrong level of detail transcribed about some factor of a finding. In a more robust system, a validation process would be necessary.

Design shortcomings

Focusing less on the user experience of a patient-facing design and more on building a soup-to-nuts prototype was a deliberate decision. Given time constraints, it was a tradeoff: Either the design could be (hopefully) more novel *without* development of a basic elastic platform, or the design could be more pedestrian *with* a stronger foundation for future work—including improved user interface design—developed instead. The "patient-facing" or part of the prototype lacks. A low-literacy conscious design, or perhaps even a full decision aid, could have been (and can now be) built on the prototype platform. So could additional tailoring options—to duration and severity of RA disease activity, for example.

Limited evidence

This significant limitation has two parts: A restriction in the kinds of findings that can be encoded today, and a cap on the quantity of evidence currently encoded. No single data schema is appropriate for all kinds of evidence, but some structure is necessary for machine readability. The schema developed so far cannot accommodate all data that might be of interest to patients. In some cases, even straightforward outcome data must be contorted to fit the schema. One particular shortcoming is that it does not accommodate arbitrary measures of statistical variance or uncertainty alongside individual findings—just 95% confidence intervals. The schema should change to support other methods, like interquartile ranges, standard deviation, and so on. It is also optimized for population-intervention-comparison-outcome data (PICO), with a weakness in the *population* dimension. The prototype eschews an existing PICO ontology (Cochrane Collaboration, n.d.-b) in favor of a simpler one that works in Google spreadsheets and is sufficient. Evidence such as treatment

algorithms, quantification of severity of contraindications and risk factors, patient anecdotes (e.g. drug reviews or video interviews), and others, require their own schemata.

The prototype is limited not only in terms of the type of evidence that it can accommodate, but also the quantity. It is obviously non-exhaustive, but there is no obvious way assess and communicate just how little evidence it contains, or what proportion of "valuable" evidence it contains or summarizes.

No feedback from patients and clinicians

Perhaps the biggest limitation of this project is that it did not involve real RA patients or caregivers, whose feedback would be necessary to guide future work on design. What is it that they find valuable about certain UIs, visualizations, or pieces of evidence? This evaluation could shift the perspective and direction of the project, or perhaps reveal that it is entirely not of value for their decision making. To a designer, it is obvious that without the human experience of people living with RA and of their caregivers and clinical care teams—those actually making decisions about these medications and observing their effects in the real world—any product intended for their use will likely fall short. What has so far been built largely represents an interpretation gained by reading literature, through the lens of personal experience as a designer and with online health resources, and projected through a philosophy and intent around a particular design artifact. It is the product of an pragmatic and intentional design experiment, rather than a comprehensive need-driven investigation.

Future work

Evaluation with target audiences

All three target audiences—patients and caregivers, clinicians and researchers, and designers and programmers—possess experience and knowledge that are invaluable to development of this work and that can only be discovered by their participation and feedback. While some decisions—such as the priority issues and kinds of data to focus on—are informed by prior research, a priority for future work is evaluation of the design, utility, and usability of the prototype and resources built with it. Among other facets, such study might include:

- Comprehensibility of specific prototype user interfaces and data visualizations.

- Following more thorough development on specific products (e.g. a decision aid) built with the prototype platform, such products' effects in shared decision making.
- Acceptability of spreadsheets as a useful for data entry, and the schemata for findings.
- Utility of technical components—for example, data visualization modules.

Some feedback can be gathered informally, such as by sharing the prototype in conversation with other designers and developers. In the event that a full decision aid were built using the platform, evaluation would have to be conducted according to medical-ethical imperatives in a more careful and rigorous way.

Design and "translation" of findings

User-facing components of the *Navigator* prototype—the overall user interface, visualizations, and so forth—are for the moment rather simplistic. As mentioned in the last section, true design investigation was limited in the work so far. Questions of user experience, data visualization, and so on, warrant deeper investigation and creative work. Some of the most obvious candidates for more work are:

- **Tailoring.** It might be possible to find evidence for specific populations—early RA, failed initial DMARD therapy, or by sex and age—that can be used to tailor *which* evidence is presented. Presentation of evidence about pain relief, for example, might be customized by allowing someone to specify their "starting point" or baseline. Then, the question would be how or whether the studied approach would be sufficient for generalization to other outcomes.
- **Much more sophisticated translation.** Overlooked outcomes—like disease progression, which is critical to the notion of why DMARDs are *necessary* in the treatment of RA—deserve attention. How can a faithful representation of evidence reflecting the imperative for disease control be "translated" into terms that matter to patients? Going further, the abandoned initial purpose of the prototype—to find a way to bring together disparate data that might together answer questions in true quality-of-life terms—could be taken up again.
- **Low-literacy, low-numeracy, and other design approaches entirely.** As the project progressed, it tended further and further away from attention to "patient-facing" design and more towards a generally useful platform for future work. One of its strengths is that the evidence in it can be presented in many different ways. The range of human needs and capacities to understand health information indicates experimentation with low-literacy design, low-numeracy design, multilingual design, finding more parallels from consumer technology and applying them, and so on. For example, there exist visualization methods for findings from multiple sources—forest plots commonly used in systematic reviews, for example. Perhaps modification of these existing techniques would work better than novel approaches? A more rigorous design process—with wider literature review—could provide fertile ground from which new data-driven designs could sprout.

More and different data

Limited data narrows the scope of any product built with the prototype platform. A deeper or wider evidence base would change its potential. One approach would be to focus exclusively on one type or suite of outcomes, such as the ACR core set, and to attempt to exhaustively encode data on that set. Another would be to enhance the prototype's capacity for *any* outcomes, including richer and more nuanced capture of measures and metrics, statistical variance, population characteristics, data quality, and so forth. Aforementioned evidence like patient-reported outcomes, treatment algorithms from clinical practice guidelines, are obvious candidates for *new* data. More historical data could be archived, or (probably more usefully) an effort could be made to focus on keeping the prototype up to date with the latest evidence. A product that might guide such a decision is one intended to keep people up to date with the latest findings on medications; the lion's share of effort for such a product would more likely be a social or community effort around encoding the findings, not a significant change to the prototype's capacity to handle data.

Elaborated harmonization and synthesis

One way to consider the presentation of heterogeneous findings from multiple sources is that it is a spectrum. At one extreme, data points are reported *as is*, in a format appropriate to the specific outcome and any idiosyncrasies. In this method, 15 findings about pain, using 15 different outcomes and measures, and gathered with 15 different methodologies, might be presented in 15 different ways, leaving the reader to interpret the meaning of all of them as a whole. In a way, that is the easiest to support technically, and potentially the worst for the end user. At the other extreme, all the data are expertly synthesized—they may be aligned to a single way of presenting them, and shown side by side, or perhaps even summarized as a single graphic or text conclusion. (Perhaps only statistically significant data would be shown.) This requires intimate knowledge of the data, advanced statistical methods, and proficiency in communicating in frames that the end user understands. It is by far the most difficult approach. Between those extremes is harmonization—the aforementioned notion of transforming or aligning findings, either statistically or just in terms of presentation, such that they are made similar enough to be applicable and understandable. The prototype currently uses a controlled vocabulary to "bring together" diverse adverse event-related outcome data under a more patient-friendly "side effects" banner, and some basic statistical and presentational harmonization of pain outcomes in a data visualization, but little else. Future work might include enhancing these capabilities, or even pursuing basic analytical synthesis capabilities—only if the data support it, and only with collaboration with people who possess the right skills.

A digital version of the RA Choice decision aid

Since the prototype began with a basic digital version of the RA Choice decision aid (Barton et al., 2014), perhaps that could be expanded to support new issues—including evidence about the issues in the *Navigator* prototype like pain and overall improvement, along with new patient-reported outcomes like sleep and fatigue (pending their quality and availability). Such a project, if the RA Choice authors were interested, might bridge the prospect of a data-driven online resource with the established value of a decision aid.

Conclusion

The project undertaken for this thesis intended to demonstrate that public engagement with medical evidence could be facilitated by building a software application with readymade and widely used development tools and design parallels from consumer Web products. Put simply, if Kayak and Google Maps can build popular, useful applications that help people derive meaning from vast troves of data, why not also in medicine? Through a research process of working with medical evidence, encoding it, designing with it, and developing with it, some answers emerged indicating significant challenges to overcome before such resources become widespread.

The unique and sensitive nature of medical data preclude its straightforward use in uncomplicated analogs to data-driven consumer Web applications. At the time of its production, evidence about rheumatoid arthritis medications is imbued with biases and essential qualities that govern its safe application —and that ultimately limit applicability in patients' lives and conclusions that may reasonably be drawn. An incomplete list of such qualities includes heterogeneity in populations studied or to whom findings may apply, methods of production, outcomes observed, study and statistical methodologies, representativeness and statistical significance, omission of data or details, limits imposed by data encoding and publishing format, and numerous sources of risk of bias. Access to such evidence—in both print and digital, machine-readable formats suitable for ingestion and use by software—is often restricted. It is not made systematically available, likely in part due to the aforementioned extreme heterogeneity. In these ways, medical data may differ from those that fuel consumer Web applications—meteorological data, road and navigation data, or flight schedule and price data, for instance. Sophisticated methods are required to finesse medical evidence for provision in a public-facing user interface to explore it. At every stage of working with the evidence, deliberate decisions shape the ultimate product—sourcing it, deciding which dimensions to encode and at what level of detail, processing it, translating it to patient-friendly terms, and ultimately displaying it. Nuance may be traded for clarity, but not without the introduction of a caveat. Each decision—like whether to articulate in detailed and machine-readable terms the details of the person to whom a finding applies—defines the ways that person can interact with the information.

This project has demonstrated that indeed there *are* tools (e.g. Google Spreadsheets and open-source software development frameworks) and tactics (e.g. risk communication visualizations and user interface designs) that can be assembled into a platform for engagement with medical evidence. However, its success is sharply limited by shortcomings that continue to emerge as data are added and the platform develops. Creating a platform for public engagement with various kinds of evidence about medications that treat rheumatoid arthritis is an aspiration without a specific endpoint, and of a scope far beyond that of an undergraduate thesis project. It demands the participation of patients and caregivers, researchers, designers, clinicians, programmers, and other specialists. Together they can discover different approaches not only to the design of such a platform, but also to future research methodologies, patient-reported outcomes, and data-publishing practices. Hopefully, such discoveries will permit a thread to be drawn from real-world needs to the production of knowledge about medications, made visible through accessible user interfaces.

Bibliography

This bibliography includes some sources not cited in this research paper, but which supply evidence used in the prototype.

1. Abadie, R., Weymiller, A. J., Tilburt, J., Shah, N. D., Charles, C., Gafni, A., & Montori, V. M. (2009). Clinician's use of the Statin Choice decision aid in patients with diabetes: a videographic study nested in a randomized trial. Journal of Evaluation in Clinical Practice, 15(3), 492–497. http://doi.org/10.1111/j.1365-2753.2008.01048.x

2. Abramson, S. (n.d.). Expected Outcomes in Rheumatoid Arthritis: An Historical Perspective (The Biologic Revolution in Rheumatoid Arthritis: How TNF Antagonists Have Optimized RA Treatment and Outcomes). Retrieved from http://www.medscape.org/viewarticle/464118_2

3. Agoritsas, T., Heen, A. F., Brandt, L., Alonso-Coello, P., Kristiansen, A., Akl, E. A., … Vandvik, P. O. (2015). Decision aids that really promote shared decision making: the pace quickens. BMJ, 350(feb10 14), g7624–g7624. http://doi.org/10.1136/bmj.g7624

4. Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., … Schünemann, H. (2011). Framing of health information messages. In Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD006777.pub2

5. Baldassari, A. R., Cleveland, R. J., Jonas, B. L., Conn, D. L., Moreland, L. W., Bridges, S. L., & Callahan, L. F. (2014). Socioeconomic Disparities in the Health of African Americans With Rheumatoid Arthritis From the Southeastern United States. Arthritis Care & Research, 66(12), 1808–1817. http://doi.org/10.1002/acr.22351

6. Barratt, A. (2008). Evidence Based Medicine and Shared Decision Making: The challenge of getting both evidence and preferences into health care. Patient Education and Counseling, 73(3), 407–412. http://doi.org/10.1016/j.pec.2008.07.054

7. Barton, J. L. (2009). Patient preferences and satisfaction in the treatment of rheumatoid arthritis with biologic therapy. Patient Preference and Adherence, 3, 335–344.

8. Barton, J. L., Koenig, C. J., Evans-Young, G., Trupin, L., Anderson, J., Ragouzeos, D., … Yelin, E. H. (2014). The design of a low literacy decision aid about rheumatoid arthritis medications developed in three languages for use during the clinical encounter. BMC Medical Informatics and Decision Making, 14(1), 104. http://doi.org/10.1186/s12911-014-0104-8

9. Blumenauer, B., Judd, M., Wells, G., Burls, A., Cranney, A., Hochberg, M., & Tugwell, P. (2002). Infliximab for the treatment of rheumatoid arthritis. The Cochrane Database of Systematic Reviews, (3), CD003785. http://doi.org/10.1002/14651858.CD003785

10. Boers, M., Kirwan, J. R., & Tugwell, P. (2014). The OMERACT Handbook. OMERACT.

11. Bolge, S. C., Goren, A., & Tandon, N. (2015). Reasons for discontinuation of subcutaneous biologic therapy in the treatment of rheumatoid arthritis: a patient perspective. Patient Preference and Adherence, 9, 121–131. http://doi.org/10.2147/PPA.S70834

12. Brenner, G. M., & Stevens, C. W. (2013). Pharmacology, 4th edition. Elsevier.

13. Busse, J. W., Bartlett, S. J., Dougados, M., Johnston, B. C., Guyatt, G. H., Kirwan, J. R., … Wells, G. A. (2015). Optimal Strategies for Reporting Pain in Clinical Trials and Systematic Reviews: Recommendations from an OMERACT 12 Workshop. The Journal of Rheumatology, jrheum.141440. http://doi.org/10.3899/jrheum.141440

14. Centers for Disease Control and Prevention. (2010). Current Depression Among Adults—United States, 2006 and 2008. Morbidity and Mortality Weekly Report, 59(38), 1229–1235.

15. Centers for Disease Control and Prevention. (2012). Chronic obstructive pulmonary disease among adults—United States, 2011. Morbidity and Mortality Weekly Report, 61(46), 938–943.

16. Centers for Disease Control and Prevention. (2014). Current Asthma Population Estimates—in thousands by Age, NHIS, 2013 (Data table). Retrieved from http://www.cdc.gov/asthma/nhis/2013/table3-1.htm#modalIdString_CDCTable_0

17. Centers for Disease Control and Prevention. (2015). Prevalence of Diagnosed and Undiagnosed HIV Infection—United States, 2008–2012. Morbidity and Mortality Weekly Report, 64(24), 657–662.

18. Charles, C., Gafni, A., & Whelan, T. (1997). Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango). Social Science & Medicine, 44(5), 681–692. http://doi.org/10.1016/S0277-9536(96)00221-3

19. Cimino, J. J. (1998). Desiderata for Controlled Medical Vocabularies in the Twenty-First Century. Methods of Information in Medicine, 37(4-5), 394–403.

20. Clark, P., Casas, E., Tugwell, P., Medina, C., Gheno, C., Tenorio, G., & Orozco, J. A. (1993). Hydroxychloroquine Compared with Placebo in Rheumatoid ArthritisA Randomized, Controlled Trial. Annals of Internal Medicine, 119(11), 1067–1071. http://doi.org/10.7326/0003-4819-119-11-199312010-00002

21. Cochrane Collaboration. (n.d.). PICO ontology | Cochrane Linked Data. Retrieved February 13, 2016, from http://linkeddata.cochrane.org/pico-ontology

22. Constantinescu, F., Goucher, S., Weinstein, A., & Fraenkel, L. (2009). Racial Disparities in Treatment Preferences for Rheumatoid Arthritis. Medical Care, 47(3), 350–355. http://doi.org/10.1097/MLR.0b013e31818af829

23. Constantinescu, F., Goucher, S., Weinstein, A., Smith, W., & Fraenkel, L. (2009). Understanding Why Rheumatoid Arthritis Patient Treatment Preferences Differ by Race. Arthritis and Rheumatism, 61(4), 413–418. http://doi.org/10.1002/art.24338

24. Crockford, D. (2006, December 6). JSON: The Fat-Free Alternative to XML. Retrieved March 1, 2016, from http://www.json.org/fatfree.html

25. Curtis, J. R., Jain, A., Askling, J., Bridges, S. L., Carmona, L., Dixon, W., … Zink, A. (2010). A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. Seminars in Arthritis and Rheumatism, 40(1), 2–14.e1. http://doi.org/10.1016/j.semarthrit.2010.03.003

26. Davidoff, F., Haynes, B., Sackett, D., & Smith, R. (1995). Evidence Based Medicine: A New Journal To Help Doctors Identify The Information They Need. BMJ: British Medical Journal, 310(6987), 1085–1086.

27. Deshpande, P. R., Rajan, S., Sudeepthi, B. L., & Abdul Nazir, C. P. (2011). Patient-reported outcomes: A new era in clinical research. Perspectives in Clinical Research, 2(4), 137–144. http://doi.org/10.4103/2229-3485.86879

28. Dutta-Bergman, M. (2003). Trusted Online Sources of Health Information: Differences in Demographics, Health Beliefs, and Health-Information Orientation. Journal of Medical Internet Research, 5(3). http://doi.org/10.2196/jmir.5.3.e21

29. Eddy, D. M. (2005). Evidence-Based Medicine: A Unified Approach. Health Affairs, 24(1), 9–17. http://doi.org/10.1377/hlthaff.24.1.9

30. Edwards, A. G., Naik, G., Ahmed, H., Elwyn, G. J., Pickles, T., Hood, K., & Playle, R. (2013). Personalised risk communication for informed decision making about taking screening tests. In The Cochrane Collaboration (Ed.), Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD001865.pub3

31. Elwyn, G., Edwards, A., Kinnersley, P., & Grol, R. (2000). Shared decision making and the concept of equipoise: the competences of involving patients in healthcare choices. The British Journal of General Practice, 50(460), 892.

32. Elwyn, G., Lloyd, A., Joseph-Williams, N., Cording, E., Thomson, R., Durand, M.-A., & Edwards, A. (2013). Option Grids: Shared decision making made easier. Patient Education and Counseling, 90(2), 207-212. http://doi.org/10.1016/j.pec.2012.06.036

33. Emanuel, E., & Emanuel, L. (1992). Four models of the physician-patient relationship. JAMA, 267(16), 2221-2226. http://doi.org/10.1001/jama.1992.03480160079038

34. Epstein, R. M., & Gramling, R. E. (2013). What Is Shared in Shared Decision Making? Complex Decisions When the Evidence Is Unclear. Medical Care Research and Review, (1 suppl), 94S-112S. http://doi.org/10.1177/1077558712459216

35. Eriksson, J. K., Neovius, M., Bratt, J., Petersson, I. F., van Vollenhoven, R. F., Geborek, P., & Ernestam, S. (2013). Biological vs. conventional combination treatment and work loss in early rheumatoid arthritis: a randomized trial. JAMA Internal Medicine, 173(15), 1407-1414. http://doi.org/10.1001/jamainternmed.2013.7801

36. Facebook. (n.d.). A JavaScript library for building user interfaces | React. Retrieved March 1, 2016, from https://facebook.github.io/react/index.html

37. Felson, D. T., Anderson, J. J., Boers, M., Bombardier, C., Chernoff, M., Fried, B., ... Lightfoot, R. (1993). The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis and Rheumatism, 36(6), 729-740.

38. Filkova, M., Carvalho, J., Norton, S., Scott, D., Mant, T., Cope, A., ... Galloway, J. (2015). Polypharmacy Is a Predictor of Hospitalisation in Patients with Rheumatoid Arthritis [abstract]. Arthritis Rheumatol., 67 (suppl 10). Retrieved from http://acrabstracts.org/abstract/polypharmacy-is-a-predictor-of-hospitalisation-in-patients-with-rheumatoid-arthritis/

39. Food and Drug Administration. (1999, February). Guidance for Industry: Clinical Development Programs for Drugs, Devices, and Biological Products for the Treatment of Rheumatoid Arthritis (RA). Food and Drug Administration. Retrieved from http://www.fda.gov/downloads/Drugs/.../Guidances/ucm071579.pdf

40. Food and Drug Administration. (2006, January). Guidance for Industry: Adverse Reactions Section of Labeling for Human Prescription Drug and Biological Products – Content and Format. Food and Drug Administration. Retrieved from http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm075057.pdf

41. Fraenkel, L., Bogardus, S., Concato, J., & Felson, D. (2001). Preference for disclosure of information among patients with rheumatoid arthritis. Arthritis and Rheumatism, 45(2), 136-139. http://doi.org/10.1002/1529-0131(200104)45:2<136::AID-ANR165>3.0.CO;2-P

42. Fraenkel, L., Bogardus, S., Concato, J., & Felson, D. (2002). Unwillingness of rheumatoid arthritis patients to risk adverse effects. Rheumatology, 41(3), 253-261. http://doi.org/10.1093/rheumatology/41.3.253

43. Fraenkel, L., Bogardus, S. T., Concato, J., Felson, D. T., & Wittink, D. R. (2004). Patient preferences for treatment of rheumatoid arthritis. Annals of the Rheumatic Diseases, 63(11), 1372-1378. http://doi.org/10.1136/ard.2003.019422

44. Fraenkel, L., Peters, E., Charpentier, P., Olsen, B., Errante, L., Schoen, R. T., & Reyna, V. (2012). A Decision Tool to Improve the Quality of Care in Rheumatoid Arthritis. Arthritis Care & Research, 64(7), 977-985. http://doi.org/10.1002/acr.21657

45. Frank Buttgereit, & Gerd-Rüdger Burmester. (2008). Glucocorticoids. In Primer on the Rheumatic Diseases, 13th edition (pp. 644-650). New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-0-387-68566-3

46. Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. Health Psychology, 28(2), 210-216. http://doi.org/10.1037/a0014474

47. Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Borden, W. B., ... Turner, M. B. (2013). Heart Disease and Stroke Statistics–2013 Update A Report From the American Heart Association. Circulation, 127(1), e6-e245. http://doi.org/10.1161/CIR.0b013e31828124ad

48. Golicki, D., Newada, M., Lis, J., Pol, K., Hermanowski, T., & Tłustochowicz, M. (2012). Leflunomide in monotherapy of rheumatoid arthritis: meta-analysis of randomized trials. Polskie Archiwum Medycyny Wewn trznej, 122(1-2), 22–32.

49. Gossec, L., Dougados, M., & Dixon, W. (2015). Patient-reported outcomes as end points in clinical trials in rheumatoid arthritis. RMD Open, 1(1), e000019. http://doi.org/10.1136/rmdopen-2014-000019

50. Greenhalgh, T. (2014). How to Read a Paper: The Basics of Evidence-Based Medicine (5 edition). Chichester, West Sussex: BMJ Books.

51. Gresh, D., Deleris, L., & Gasparini, L. (2012). Visualizing Risk. IBM Research.

52. Guyatt, G., Cairns, J., & Churchill, D. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. JAMA, 268(17), 2420–2425. http://doi.org/10.1001/jama.1992.03490170092032

53. Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ, 336(7650), 924–926. http://doi.org/10.1136/bmj.39489.470347.AD

54. Guyatt, G., Rennie, D., Meade, M., & Cook, D. (2008). Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice, Second Edition (2 edition). New York: McGraw-Hill Professional.

55. Guyatt, G., & Vandvik, P. O. (2013). Creating clinical practice guidelines: Problems and solutions. Chest, 144(2), 365–367. http://doi.org/10.1378/chest.13-0463

56. Hawkins, R. P., Kreuter, M., Resnicow, K., Fishbein, M., & Dijkstra, A. (2008). Understanding tailoring in communicating about health. Health Education Research, 23(3), 454–466. http://doi.org/10.1093/her/cyn004

57. Haynes, R. B., Ackloo, E., Sahota, N., McDonald, H. P., & Yao, X. (1996). Interventions for enhancing medication adherence. In Cochrane Database of Systematic Reviews. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com.proxy.lib.uwaterloo.ca/doi/10.1002/14651858.CD000011.pub3/abstract

58. Heisler, M., Vijan, S., Anderson, R. M., Ubel, P. A., Bernstein, S. J., & Hofer, T. P. (2003). When Do Patients and Their Physicians Agree on Diabetes Treatment Goals and Strategies, and What Difference Does It Make? Journal of General Internal Medicine, 18(11), 893–902. http://doi.org/10.1046/j.1525-1497.2003.21132.x

59. Helmick, C. G., Felson, D. T., Lawrence, R. C., Gabriel, S., Hirsch, R., Kwoh, C. K., … National Arthritis Data Workgroup. (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. Arthritis & Rheumatism, 58(1), 15–25. http://doi.org/10.1002/art.23177

60. Hesse, B. W., Nelson, D. E., Kreps, G. L., Croyle, R. T., Arora, N. K., Rimer, B. K., & Viswanath, K. (2005). Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Archives of Internal Medicine, 165(22), 2618–2624. http://doi.org/10.1001/archinte.165.22.2618

61. Hoffman, A. S., Volk, R. J., Saarimaki, A., Stirling, C., Li, L. C., Härter, M., … Llewellyn-Thomas, H. (2013). Delivering patient decision aids on the Internet: definitions, theories, current evidence, and emerging research areas. BMC Medical Informatics and Decision Making, 13(Suppl 2), S13. http://doi.org/10.1186/1472-6947-13-S2-S13

62. Holmström, I., & Röing, M. (2010). The relation between patient-centeredness and patient empowerment: A discussion on concepts. Patient Education and Counseling, 79(2), 167–172. http://doi.org/10.1016/j.pec.2009.08.008

63. Horne, R., & Weinman, J. (1999). Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. Journal of Psychosomatic Research, 47(6), 555–567. http://doi.org/10.1016/S0022-3999(99)00057-4

64. ICSI. (2013). Adult Depression in Primary Care Guideline. Retrieved from https://www.icsi.org/guidelines_more/catalog_guidelines_and_more/catalog_guidelines/catalog_behavioral_health_guidelines/depression/

65. Institute of Medicine. (2001). Crossing the Quality Chasm: A New Health System for the 21st Century. Retrieved from http://www.iom.edu/Reports/2001/Crossing-the-Quality-Chasm-A-New-Health-System-for-the-21st-Century.aspx

66. Jackson, T., Wang, Y., Wang, Y., & Fan, H. (2014). Self-Efficacy and Chronic Pain Outcomes: A Meta-Analytic Review. The Journal of Pain, 15(8), 800–814. http://doi.org/10.1016/j.jpain.2014.05.002

67. Jansen, J. P., Buckley, F., Dejonckheere, F., & Ogale, S. (2014). Comparative efficacy of biologics as monotherapy and in combination with methotrexate on patient reported outcomes (PROs) in rheumatoid arthritis patients with an inadequate response to conventional DMARDs - a systematic review and network meta-analysis. Health and Quality of Life Outcomes, 12, 102. http://doi.org/10.1186/1477-7525-12-102

68. Katchamart, W., Trudeau, J., Phumethum, V., & Bombardier, C. (2010). Methotrexate monotherapy versus methotrexate combination therapy with non-biologic disease modifying anti-rheumatic drugs for rheumatoid arthritis. In Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD008495

69. Kirwan, J. R., Hewlett, S. E., Heiberg, T., Hughes, R. A., Carr, M., Hehir, M., … Wale, J. (2005). Incorporating the patient perspective into outcome assessment in rheumatoid arthritis--progress at OMERACT 7. The Journal of Rheumatology, 32(11), 2250–2256.

70. Kirwan, J. R., Newman, S. P., Tugwell, P., & Wells, G. (2009). Patient perspective on outcomes in rheumatology -- a position paper for OMERACT 9. The Journal of Rheumatology, 36(9), 2067-70. http://doi.org/10.3899/jrheum.090359

71. Klippel, J. H., Stone, J. H., Crofford, L. J., & White, P. H. (Eds.). (2008). Primer on the Rheumatic Diseases, 13th edition. New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-0-387-68566-3

72. Kon AA. (2010). The shared decision-making continuum. JAMA, 304(8), 903–904. http://doi.org/10.1001/jama.2010.1208

73. Kreuter, M. W., Oswald, D. L., Bull, F. C., & Clark, E. M. (2000). Are tailored health education materials always more effective than non-tailored materials? Health Education Research, 15(3), 305–315. http://doi.org/10.1093/her/15.3.305

74. Kristiansen, A., Brandt, L., Alonso-Coello, P., Agoritsas, T., Akl, E. A., Conboy, T., … Vandvik, P. O. (2014). Development of a novel multilayered presentation format for clinical practice guidelines. Chest. http://doi.org/10.1378/chest.14-1366

75. Kvien, T. (2004). Epidemiology and burden of illness of rheumatoid arthritis. PharmacoEconomics, 22, 1–12.

76. Lawrence, R. C., Felson, D. T., Helmick, C. G., Arnold, L. M., Choi, H., Deyo, R. A., … National Arthritis Data Workgroup. (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. Arthritis and Rheumatism, 58(1), 26–35. http://doi.org/10.1002/art.23176

77. LeBlanc A, Herrin J, Williams MD, & et al. (2015). Shared decision making for antidepressants in primary care: A cluster randomized trial. JAMA Internal Medicine, 1–10. http://doi.org/10.1001/jamainternmed.2015.5214

78. Leppin, A. L., Montori, V. M., & Gionfriddo, M. R. (2015). Minimally Disruptive Medicine: A Pragmatically Comprehensive Model for Delivering Care to Patients with Multiple Chronic Conditions. Healthcare, 3(1), 50–63. http://doi.org/10.3390/healthcare3010050

79. Lethaby, A., Lopez-Olivo, M. A., Maxwell, L., Burls, A., Tugwell, P., & Wells, G. A. (2013). Etanercept for the treatment of rheumatoid arthritis. The Cochrane Database of Systematic Reviews, 5, CD004525. http://doi.org/10.1002/14651858.CD004525.pub2

80. Lin, G. A., & Fagerlin, A. (2014). Shared Decision Making State of the Science. Circulation: Cardiovascular Quality and Outcomes, 7(2), 328–334. http://doi.org/10.1161/CIRCOUTCOMES.113.000322

81. Lopez-Olivo, M. A., Amezaga Urruela, M., McGahan, L., Pollono, E. N., & Suarez-Almazor, M. E. (2015). Rituximab for rheumatoid arthritis. In Cochrane Database of Systematic Reviews.

John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com.proxy.lib.uwaterloo.ca/doi/10.1002/14651858.CD007356.pub2/abstract

82. Lopez-Olivo, M. A., Siddhanamatha, H. R., Shea, B., Tugwell, P., Wells, G. A., & Suarez-Almazor, M. E. (2014). Methotrexate for treating rheumatoid arthritis. In Cochrane Database of Systematic Reviews. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD000957.pub2/abstract

83. Lundgren, R. E. (2013). Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks (5 edition). Hoboken, New Jersey: Wiley-IEEE Press.

84. Matteson, E. L. (2001). Rheumatoid Arthritis Treatment. In Primer on the Rheumatic Diseases, 12th edition (pp. 225–231).

85. Maxwell, L., & Singh, J. A. (2009). Abatacept for rheumatoid arthritis. In Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD007277.pub2

86. Mayo Clinic. (n.d.-a). Etanercept (Subcutaneous Route) Description and Brand Names - Mayo Clinic. Retrieved February 8, 2016, from http://www.mayoclinic.org/drugs-supplements/etanercept-subcutaneous-route/description/drg-20066850

87. Mayo Clinic. (n.d.-b). Rheumatoid arthritis Treatments and drugs - Mayo Clinic. Retrieved February 8, 2016, from http://www.mayoclinic.org/diseases-conditions/rheumatoid-arthritis/basics/treatment/con-20014868

88. Mayo Clinic Shared Decision Making National Resource Center. (n.d.-a). Depression Medication Choice. Retrieved February 8, 2016, from http://shareddecisions.mayoclinic.org/decision-aid-information/decision-aids-for-chronic-disease/depression-medication-choice/

89. Mayo Clinic Shared Decision Making National Resource Center. (n.d.-b). Diabetes Medication Choice. Retrieved February 8, 2016, from https://diabetesdecisionaid.mayoclinic.org/

90. Mayo Clinic Shared Decision Making National Resource Center. (n.d.-c). Statin Choice Decision Aid. Retrieved February 8, 2016, from https://statindecisionaid.mayoclinic.org/

91. Mertens, M., & Singh, J. (2008). Anakinra for rheumatoid arthritis. In Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD005121.pub2

92. Molewijk, B., Stiggelbout, A. M., Otten, W., Dupuis, H. M., & Kievit, J. (2008). First the facts, then the values? Implicit normativity in evidence-based decision aids for shared decision-making. Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen, 102(7), 415–420. http://doi.org/10.1016/j.zefq.2008.08.014

93. Montori, V. M., Breslin, M., Maleska, M., & Weymiller, A. J. (2007). Creating a Conversation: Insights from the Development of a Decision Aid. PLoS Med, 4(8), e233. http://doi.org/10.1371/journal.pmed.0040233

94. Montori, V. M., LeBlanc, A., Buchholz, A., Stilwell, D. L., & Tsapas, A. (2013). Basing information on comprehensive, critically appraised, and up-to-date syntheses of the scientific evidence: a quality dimension of the International Patient Decision Aid Standards. BMC Medical Informatics and Decision Making, 13(Suppl 2), S5. http://doi.org/10.1186/1472-6947-13-S2-S5

95. Navarro-Sarabia, F., Ariza-Ariza, R., Hernandez-Cruz, B., & Villanueva, I. (2005). Adalimumab for treating rheumatoid arthritis. In Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD005113.pub2

96. Nota, I., Drossaert, C. H. C., Taal, E., Vonkeman, H. E., & van de Laar, M. A. F. J. (2014). Patient participation in decisions about disease modifying anti-rheumatic drugs: a cross-sectional survey. BMC Musculoskeletal Disorders, 15, 333. http://doi.org/10.1186/1471-2474-15-333

97. Nwankwo, T., Yoon, S. S., Burt, V., & Gu, Q. (2013). Hypertension among adults in the United States: National Health and Nutrition Examination Survey, 2011-2012. NCHS Data Brief, (133), 1–8.

98. Oliver, A. M., & St Clair, E. W. (2008). Rheumatoid Arthritis Treatment and Assessment. In Primer on the Rheumatic Diseases, 13th edition (pp. 133–141). New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-0-387-68566-3

99. Olofsson, T., Petersson, I. F., Eriksson, J. K., Englund, M., Simard, J. F., Nilsson, J.-Å., … Neovius, M. (2014). Predictors of work disability during the first 3 years after diagnosis in a national rheumatoid arthritis inception cohort. Annals of the Rheumatic Diseases, 73(5), 845–853. http://doi.org/10.1136/annrheumdis-2012-202911

100. Osiri, M., Shea, B., Robinson, V., Suarez-Almazor, M., Strand, V., Tugwell, P., & Wells, G. (2003). Leflunomide for the treatment of rheumatoid arthritis: a systematic review and metaanalysis. The Journal of Rheumatology, 30(6), 1182–1190.

101. Pannucci, C. J., & Wilkins, E. G. (2010). Identifying and Avoiding Bias in Research. Plastic and Reconstructive Surgery, 126(2), 619–625. http://doi.org/10.1097/PRS.0b013e3181de24bc

102. Perneger, T. V., & Agoritsas, T. (2011). Doctors and patients' susceptibility to framing bias: a randomized trial. Journal of General Internal Medicine, 26(12), 1411–1417. http://doi.org/10.1007/s11606-011-1810-x

103. Petrescu, P. (2014, October 1). Google Organic Click-Through Rates in 2014. Retrieved February 8, 2016, from https://moz.com/blog/google-organic-click-through-rates-in-2014

104. Pincus, T. (2005). The American College of Rheumatology (ACR) Core Data Set and derivative "patient only" indices to assess rheumatoid arthritis. Clinical and Experimental Rheumatology, 23(5 Suppl 39), S109–113.

105. Puolakka, K., Kautiainen, H., Möttönen, T., Hannonen, P., Korpela, M., Hakala, M., … Leirisalo-Repo, M. (2005). Early suppression of disease activity is essential for maintenance of work capacity in patients with recent-onset rheumatoid arthritis: Five-year experience from the FIN-RACo trial. Arthritis & Rheumatism, 52(1), 36–41. http://doi.org/10.1002/art.20716

106. Rader, T., Maxwell, L., Ghogomu, E., Tugwell, P., & Welch, V. (2011, February). Should I take methotrexate (Rheumatrex®) alone or with other disease- modifying anti-rheumatic drugs for rheumatoid arthritis? A Cochrane decision aid to discuss options with your doctor. Cochrane Musculoskeletal Group. Retrieved from http://musculoskeletal.cochrane.org/sites/musculoskeletal.cochrane.org/files/uploads/Should%20I%20take%20methotrexate-21%20Mar%202011.pdf

107. Rauscher, V., Englbrecht, M., van der Heijde, D., Schett, G., & Hueber, A. J. (2015). High Degree of Nonadherence to Oral Disease-modifying Antirheumatic Drugs in Patients with Rheumatoid Arthritis. The Journal of Rheumatology. http://doi.org/10.3899/jrheum.140982

108. Rendas-Baum, R., Bayliss, M., Kosinski, M., Raju, A., Zwillich, S. H., Wallenstein, G. V., & Koncz, T. (2014). Measuring the effect of therapy in rheumatoid arthritis clinical trials from the patient's perspective. Current Medical Research and Opinion, 30(7), 1391–1403. http://doi.org/10.1185/03007995.2014.896328

109. Richards, B. L., Whittle, S. L., & Buchbinder, R. (2011). Antidepressants for pain management in rheumatoid arthritis. In The Cochrane Collaboration (Ed.), Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD008920.pub2

110. Richards, J. S., Dowell, S. M., Quinones, M. E., & Kerr, G. S. (2015). How to use biologic agents in patients with rheumatoid arthritis who have comorbid disease. The BMJ, 351, h3658. http://doi.org/10.1136/bmj.h3658

111. Roth, F. (2012). Visualizing Risk: The Use of Graphical Elements in Risk Analysis and Communications.

112. Ruiz Garcia, V., Jobanputra, P., Burls, A., Cabello, J. B., Vela Casasempere, P., Bort-Marti, S., & Kynaston-Pearson, F. J. B. (2014). Certolizumab pegol (CDP870) for rheumatoid arthritis in adults. The Cochrane Database of Systematic Reviews, 9, CD007649. http://doi.org/10.1002/14651858.CD007649.pub3

113. Sanderson, T., Morris, M., Calnan, M., Richards, P., & Hewlett, S. (2010). Patient perspective of measuring treatment efficacy: the Rheumatoid Arthritis Patient Priorities for Pharmacological Interventions (RAPP-PI) outcomes. Arthritis Care & Research, 62(5), 647–656. http://doi.org/10.1002/acr.20151

114. Shaw, D., & Elger, B. (2013). Evidence-based persuasion: An ethical imperative. JAMA, 309(16), 1689–1690. http://doi.org/10.1001/jama.2013.2179

115. Shay, L. A., & Lafata, J. E. (2015). Where is the evidence? A systematic review of shared decision making and patient outcomes. Medical Decision Making: An International Journal of the Society for Medical Decision Making, 35(1), 114–131. http://doi.org/10.1177/0272989X14551638

116. Singh, J. A., Beg, S., & Lopez-Olivo, M. A. (2010). Tocilizumab for rheumatoid arthritis. In Cochrane Database of Systematic Reviews. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com.proxy.lib.uwaterloo.ca/doi/10.1002/14651858.CD008331.pub2/abstract

117. Singh, J. A., Christensen, R., Wells, G. A., Suarez-Almazor, M. E., Buchbinder, R., Lopez-Olivo, M. A., … Tugwell, P. (2009). Biologics for rheumatoid arthritis: an overview of Cochrane reviews. In Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD007848.pub2

118. Singh, J. A., Noorbaloochi, S., & Singh, G. (2010). Golimumab for rheumatoid arthritis: a systematic review. The Journal of Rheumatology, 37(6), 1096–1104. http://doi.org/10.3899/jrheum.091466

119. Singh, J. A., Saag, K. G., Bridges, S. L., Akl, E. A., Bannuru, R. R., Sullivan, M. C., … McAlindon, T. (2016). 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. Arthritis & Rheumatology (Hoboken, N.J.), 68(1), 1–26. http://doi.org/10.1002/art.39480

120. Smolen, J. S., Kalden, J. R., Scott, D. L., Rozman, B., Kvien, T. K., Larsen, A., … Rosenburg, R. (1999). Efficacy and safety of leflunomide compared with placebo and sulphasalazine in active rheumatoid arthritis: a double-blind, randomised, multicentre trial. The Lancet, 353(9149), 259–266. http://doi.org/10.1016/S0140-6736(98)09403-3

121. Smolen, J. S., Landewé, R., Breedveld, F. C., Dougados, M., Emery, P., Gaujoux-Viala, C., … van der Heijde, D. (2013). EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Annals of the Rheumatic Diseases, 69(6), 964–975. http://doi.org/10.1136/ard.2009.126532

122. Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing Uncertainty About the Future. Science, 333(6048), 1393–1400. http://doi.org/10.1126/science.1191181

123. Stacey, D., Légaré, F., Col, N. F., Bennett, C. L., Barry, M. J., Eden, K. B., … Wu, J. H. (2012). Decision aids for people facing health treatment or screening decisions. In Cochrane Database of Systematic Reviews. John Wiley & Sons, Ltd. Retrieved from http://onlinelibrary.wiley.com.proxy.lib.uwaterloo.ca/doi/10.1002/14651858.CD001431.pub4/abstract

124. Stamer, M., Schmacke, N., & Richter, P. (2013). Noncompliance: A Never-Ending Story. Understanding the Perspective of Patients with Rheumatoid Arthritis. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 14(3). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1932

125. Steiman, A. J., Pope, J. E., Thiessen-Philbrook, H., Li, L., Barnabe, C., Kalache, F., … Bykerk, V. (2013). Non-biologic disease-modifying antirheumatic drugs (DMARDs) improve pain in inflammatory arthritis (IA): a systematic literature review of randomized controlled trials. Rheumatology International, 33(5), 1105–1120. http://doi.org/10.1007/s00296-012-2619-6

126. Stiggelbout, A. M., Weijden, T. V. d., Wit, M. P. T. D., Frosch, D., Legare, F., Montori, V. M., … Elwyn, G. (2012). Shared decision making: really putting patients at the centre of healthcare. BMJ, 344(jan27 1), e256–e256. http://doi.org/10.1136/bmj.e256

127. Strand, V., Burmester, G. R., Zerbini, C. A. F., Mebus, C. A., Zwillich, S. H., Gruben, D., & Wallenstein, G. V. (2014). Tofacitinib with methotrexate in third-line treatment of patients with active rheumatoid arthritis: Patient-reported outcomes from a Phase 3 trial. Arthritis Care & Research. http://doi.org/10.1002/acr.22453

128. Strand V, Cohen S, Schiff M, & et al. (1999). Treatment of active rheumatoid arthritis with leflunomide compared with placebo and methotrexate. Archives of Internal Medicine, 159(21), 2542–2550. http://doi.org/10.1001/archinte.159.21.2542

129. Suarez-Almazor, M. E., Belseck, E., Shea, B., Homik, J., Wells, G., & Tugwell, P. (2000). Antimalarials for treating rheumatoid arthritis. The Cochrane Database of Systematic Reviews, (4), CD000959. http://doi.org/10.1002/14651858.CD000959

130. Suarez-Almazor, M. E., Belseck, E., Shea, B., Tugwell, P., & Wells, G. A. (1998). Sulfasalazine for treating rheumatoid arthritis. In The Cochrane Collaboration (Ed.), Cochrane Database of Systematic Reviews. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/14651858.CD000958

131. Tehlirian, C. V., & Bathon, J. M. (2008). Rheumatoid Arthritis Clinical and Laboratory Manifestations. In Primer on the Rheumatic Diseases, 13th edition (pp. 133–141). New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-0-387-68566-3

132. Treharne, G. J., Douglas, K. M. J., Iwaszko, J., Panoulas, V. F., Hale, E. D., Mitton, D. L., … Kitas, G. D. (2007). Polypharmacy among people with rheumatoid arthritis: the role of age, disease duration and comorbidity. Musculoskeletal Care, 5(4), 175–190. http://doi.org/10.1002/msc.112

133. Trevena, L. J., Zikmund-Fisher, B. J., Edwards, A., Gaissmaier, W., Galesic, M., Han, P. K., … Woloshin, S. (2013). Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Medical Informatics and Decision Making, 13(Suppl 2), S7. http://doi.org/10.1186/1472-6947-13-S2-S7

134. Vandvik, P. O. (2015, July). Want to SHARE-IT? Testing Decision Aids Linked to Evidence Summaries and Guidelines for use in Clinical Consultations. Presented at the ISDM-ISEHC 2015, Sydney. Retrieved from http://www.ebhcconference.org/previous_editions/2015/presentations/WPC1_Vandvik_P.pdf

135. Vandvik, P. O., Brandt, L., Alonso-Coello, P., Treweek, S., Akl, E. A., Kristiansen, A., … Guyatt, G. (2013). Creating clinical practice guidelines we can trust, use, and share: a new era is imminent. Chest, 144(2), 381–389. http://doi.org/10.1378/chest.13-0746

136. van Vilsteren, M., Boot, C. R., Knol, D. L., van Schaardenburg, D., Voskuyl, A. E., Steenbeek, R., & Anema, J. R. (2015). Productivity at work and quality of life in patients with rheumatoid arthritis. BMC Musculoskeletal Disorders, 16. http://doi.org/10.1186/s12891-015-0562-x

137. Waldburger, J.-M., & Firestein, G. S. (2008). Rheumatoid Arthritis Epidemiology, Pathology, and Pathogenesis. In Primer on the Rheumatic Diseases, 13th edition (pp. 133–141). New York, NY: Springer New York. Retrieved from http://link.springer.com/10.1007/978-0-387-68566-3

138. Ward, B. W., Schiller, J. S., & Goodman, R. A. (2014). Multiple Chronic Conditions Among US Adults: A 2012 Update. Preventing Chronic Disease, 11. http://doi.org/10.5888/pcd11.130389

139. Wee, M. M. ter, Lems, W. F., Usan, H., Gulpen, A., & Boonen, A. (2012). The effect of biological agents on work participation in rheumatoid arthritis patients: a systematic review. Annals of the Rheumatic Diseases, 71(2), 161–171. http://doi.org/10.1136/ard.2011.154583

140. Weijden, T. van der, Légaré, F., Boivin, A., Burgers, J. S., Veenendaal, H. van, Stiggelbout, A. M., … Elwyn, G. (2010). How to integrate individual patient values and preferences in clinical practice guidelines? A research protocol. Implementation Science, 5(1), 10. http://doi.org/10.1186/1748-5908-5-10

141. Williams, H. J., Willkens, R. F., Samuelson, C. O., Alarcón, G. S., Guttadauria, M., Yarboro, C., … Ward, J. R. (1985). Comparison of low-dose oral pulse methotrexate and placebo in the treatment of rheumatoid arthritis. A Controlled Clinical Trial. Arthritis & Rheumatism, 28(7), 721–730. http://doi.org/10.1002/art.1780280702

142. Wolfe, F., & Michaud, K. (2007). Resistance of rheumatoid arthritis patients to changing therapy: discordance between disease activity and patients' treatment choices. Arthritis and Rheumatism, 56(7), 2135–2142. http://doi.org/10.1002/art.22719