Published in final edited form as:

Rheum Dis Clin North Am. 2019 May ; 45(2): 275–289. doi:10.1016/j.rdc.2019.01.010.

Tools and Methods for Real-World Evidence Generation: Pragmatic Trials, Electronic Consent, and Data Linkages

Jeffrey R Curtis, MD MS MPH¹, P. Jeff Foster¹, and Kenneth G. Saag, MD MSc¹

¹Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL

Keywords

pragmatic trials; real world evidence; rheumatoid arthritis; osteoporosis; data linkage; informed consent

Introduction

Real world evidence on clinical effectiveness has assumed increasingly greater importance over time in the practice of medicine. Given the recognition that a majority of what is done in clinical practice is not supported by direct evidence, clinicians, researchers, policy makers, regulatory agencies, and other stakeholders are turning their attention to evidence that has better generalizability to inform the practical decisions that they must make day-to-day. Real world data that is used for evidence generation often refers to information about health care derived from multiple sources that are outside the typical confines of research (e.g. randomized controlled trials [RCT] of highly selected individuals) (1).

The hoped-for benefits of generating evidence and obtaining it from routine care settings, as may be recorded in electronic medical record (EMR) systems, administrative claims data (e.g. pharmacy refill data), mobile health (mHealth), and other diverse sources, is not only to improve the generalizability of research, but also to expand the available evidence for patients with significant disease heterogeneity, multi-morbidity, and more severe disease manifestations than typically would be permitted in a RCT. For example, the proportion of rheumatoid arthritis (RA) patients seen in routine care settings that would be eligible for a typical phase 3 RCT has been estimated to be 20% or even lower (2, 3). Additional factors that may affect clinical outcomes that would be irrelevant, or considered as ‘nuisance factors’ in an RCT, can also be studied more effectively using real-world data. For example, aspects of the healthcare delivery system, clinician and provider-related factors (e.g. primary care vs. specialty providers), and patient-related influences including infrastructure (e.g.

Corresponding author: Jeffrey Curtis, 510 20th Street South, Faculty Office Towers 802, Birmingham AL 35294, jcurtis@uab.edu.

Disclosures: none

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

local care pathways and treatment patterns, on-site infusion capabilities), accessibility (e.g. distance to clinical site, appointment wait times), socioeconomic status, specific comorbidities, and medication adherence all can affect outcomes. These can be studied as components of real-world evidence generation and thus improve the generalizability of research findings.

A common misconception is that real world evidence (RWE) requires researchers to stay firmly grounded in the demesne of observational studies. However, RWE is compatible with the ability to randomize, and deliver an intervention, if that is needed. Randomization can be at the individual patient level, or at a group or site level (e.g. cluster randomization) (4), which may facilitate practical study implementation and help avoid confounding related to inadvertent provider or site-specific influences on the study intervention that might apply to all patients treated at that site. The 21st Century Cures Act enabled the Food and Drug Administration to include RWE in the regulatory process, which may encompass activities including drug label expansions and post-marketing surveillance programs (5). For all the reasons above and more, RWE presents the unprecedented opportunity to extend the research enterprise beyond the typical bounds of academia and to optimally create a true learning health care system whereby research is seamlessly integrated into clinical care. However, to do so effectively, it will require new tools and approaches for evidence generation. This article will be informed by exemplar case studies in rheumatology, with a particular focus on the design features of a RCT of the live zoster vaccine in biologic-treated patients that illustrate use of new tools and methods including pragmatic trials, electronic patient consent, and data linkages.

Pragmatic Trials

Explanatory vs. Pragmatic Trials

Historically, most randomized controlled clinical trials have been explanatory in nature. The focus of such trials is efficacy i.e. whether the intervention (e.g. drug, treatment strategy, medical procedure) works under ideal conditions. These studies aimed at confirming that the desired therapeutics work as well in humans as early studies *in vitro* and in animals might suggest, are critical for regular approval of drugs, devices and biologics. However, patients tend to be highly selected, and those with any anticipated impediments to full adherence to the study protocol are excluded at the outset. The study is monitored closely for adherence to the protocol, and endpoints may be biologic proxies for the outcome (e.g. changes in a biomarker or lab test) or process measures (e.g. delivery of care, or a specified clinical action taken, at the prescribed time). However, often these trials can only provide indirect evidence for clinicians, patients, and policymakers who need this data to make decisions. In contrast, pragmatic trials focus on effectiveness (i.e. does the intervention work in normal practice?) Inclusion criteria in pragmatic trials are often less restrictive, and an ‘all-comers’ approach that is representative of the patients seen in real-world settings is preferred. Optimally, the investigators should be closer to routine practitioners and less likely to be large research shops where many phase 3 studies are conducted. Interventions are applied as they would be in routine clinical care, and as such, these trials are intended to generate direct evidence about the population to which the intervention will be ultimately applied. In contrast to

explanatory studies these studies typically trade-off better generalizability for lower internal validity. Features of explanatory vs. pragmatic trials have been summarized (6) and described along a continuum known as the Pragmatic-Explanatory Continuum Indicator Summary (7) (Figure 1).

Features that facilitate the conduct of a pragmatic trial include study outcomes that are simpler and more objectively defined (e.g. a discrete event such as a well-defined infection, fracture, myocardial infarction, or death), an intervention that can be delivered efficiently in a relatively simple fashion, and a pool of motivated participants (including both patients and providers) that are willing to participate.

A Practical Example: the VERVE trial

While a variety of pragmatic trials have been conducted in medicine (8–10), and in rheumatology (11), the design features of the VaricEcella zoster VaccinE (VERVE) trial illustrate many of the relevant concepts of a pragmatic trial (12). VERVE is a randomized controlled blinded trial of patients aged 50 and older receiving anti-tumor necrosis factor (TNF) therapy who are randomized to receive the live zoster vaccine or placebo (saline). VERVE has very simple inclusion criteria, and unless someone is being actively treated for malignancy, has a meaningful other source of immunosuppression (e.g. transplant, high dose glucocorticoid use), previously received the live zoster vaccine, or could become pregnant, all patients on any of the five TNF inhibitor (TNFi) therapies are eligible to participate. There are no disease-specific requirements (e.g. rheumatoid arthritis, psoriasis, inflammatory bowel disease), and even off-label use of TNFi treatment is permitted (e.g. sarcoidosis). A key motivating concern that has tempered enthusiasm to use this attenuated, live virus vaccine for patients receiving biologic treatments is the potential for the vaccine to cause a weakened form of the very infection (i.e. varicella zoster, manifested as chickenpox) that it is intended to reduce complications for. For that reason, safety is a main outcome of the study.

If patients were to develop vaccine-strain varicella infection related to the vaccine, it is expected to occur early (e.g. within 4 weeks). Therefore, VERVE has its second and only follow-up visit at week 6 (to allow for a comfortable margin past 4 weeks). At this visit, as at the baseline visit, biospecimens are collected for the trial to assess the trials' immunologic-based outcomes (e.g. glycoprotein enzyme-linked immunosorbent assay (gpELISA), and changes in cytokines measured via peripheral blood mononuclear cells [PBMCs]). Were a patient to develop a rash suggestive of varicella infection within 6 weeks of vaccine (or placebo) administration, skin swabs would be taken and polymerase chain reaction (PCR) performed to verify varicella infection, and subtype as vaccine-strain vs. wild-type strain.

Because the trial needed large numbers of patients to meet its objectives (e.g. > 600), an efficient process to both screen for, and recruit patients, is essential. Based on the trials' relatively simply inclusion criteria, providers and healthcare systems that can search their EMR or electronic data warehouse for patients based on demographics and medication use can easily pre-identify eligible patients. This electronic screening using tools such as i2b2 (13, 14) can generate a list of eligible individuals, and enable a study coordinator to reach

out to them (e.g. patient portal, telephone) prior to their next upcoming visit, and/or make contact with them in person at the time of their next routine appointment. This approach provides tremendous recruitment efficiency compared to the typical method of having a study coordinator passively wait for a clinician to refer patients individually to the trial.

As one final aspect of the potential efficiencies of a pragmatic trial, institutional review board approval can be streamlined. A majority (23 of 33) of the VERVE trial sites were governed by a single IRB. The new NIH guidance that is intended to streamline the IRB process and shorten time to trial activation will require that even multi-site trials use only a single IRB for approval (15). The SMART IRB platform provides a flexible, efficient enhanced version that extends capacities previously served by traditional reliance agreements, whereby which only one IRB governs the conduct of the study, and all other sites (including academic medical centers) will rely on the single, central IRB for oversight. Both academic medical center and commercial IRBs (e.g. Advarra, Western IRB) can serve to provide oversight for SMART-IRB approved studies.

Electronic Screening and Consent

A variety of barriers exists to a general clinician's participation in pragmatic clinical trials (16–19). As one of those barriers, the time commitment and feasibility of not only screening for eligible patients, but also the process of obtaining informed consent from a patient, can be burdensome in a busy clinical setting not routinely devoted to clinical trials. The details of the trial must be explained to the patient in a clear fashion, and specific elements of the study as described in the informed consent document approved by an institutional review board (IRB) must be presented. While the IRB(s) can provide well-defined guidance about what an informed consent document must contain and undergoes careful scrutiny by the IRB itself, the often informal communication of the study's nature, objectives, and details may be highly variable when communicated by a study investigator or research coordinator to a patient. To reduce variability in the understanding of the VERVE trial's key elements, a tablet-based (iPad) interactive system was developed and deployed to explain the study and facilitate obtaining informed consent. The patient is asked a series of short questions related to the trial's few inclusion criteria (e.g. current use of any of the give anti-TNF therapies, current treatment for malignancy, cessation of menstruation for more than a year [to avoid enrolling women who might become pregnant], and prior receipt of the zoster vaccine) to affirm eligibility. Patients are then presented a custom 6–7-minute video (Figure 2) developed specifically for the study that explains its details in clear and simple terms. The study coordinator (who in a PCT may be a clinic staff person less well versed in research) need not be even in the room as the video is viewed.

After the patient has watched the required video, the informed consent document is presented on the tablet, and the patient can depress the screen to obtain more information and a dictionary definition for any of the specific terms mentioned in the consent document. Having read the consent document, the patient is asked a series of questions to confirm that they understand the essential elements of the trial. If a patient does not answer correctly, and assuming the consenting clinicians believe obtained true informed consent for this patient is indeed possible, the knowledge gap can sometimes be remediated, and the potential

participant hyperlinked back to the relevant section of the informed consent document to refer back to the appropriate information in the informed consent document (Figure 3). Having pilot tested this assessment process during the development of a different pragmatic trial focused on bisphosphonate drug holidays (20), a focus group of IRB members as well as the IRBs governing the VERVE trial generally found it highly acceptable, particularly compared to the typical requirement that is generally satisfied (from a documentation standpoint) by a patient's initials appearing on the bottom of every page of the informed consent, and their signature at the end. The consent process can also entail any number of revisions or customizations required by a local IRB, or state and local regulations. For example, the state of California has specific requirements for research participation (21) (22), including the requirement to present to potential patients the text of the "experimental subject's bill of rights" which was made available on the tablet to the California participants considering the trial. Finally, patients are afforded the opportunity to ask questions of the research coordinator and site investigator, who provides an electronic countersignature to confirm enrollment.

Overall, the process of using a tablet computer to obtain informed consent from patients for a pragmatic trial was found to be highly acceptable based on experience from a pilot study (23), and site staff and patients generally were at least as comfortable, if not preferred, the tablet-based method compared to its traditional paper equivalent. In the conduct of the VERVE trial, and based on 613 patients randomized through November 2018, only 7.9% of patients who initiated the electronic consent process (and presumably had been deemed tentatively eligible for the study) answered the screening questions in a way that precluded their participation, and only 13% of people who started the electronic consent process elected to not participate. Overall, the electronic consent process was deemed acceptable and feasible by VERVE trial sites, although one of the 33 VERVE trial sites was initially not permitted by their local IRB to participate because consent was obtained electronically. The site later reversed its position, perhaps facilitated by the eConsent guidance from the FDA and the Department of Health and Human Services that was carefully adopted for the VERVE trial (24).

Once patients decide to participate, their signature is collected electronically on the tablet (i.e. with their finger or stylus). This feature, coupled with the few patient-facing screening questions and short data collection elements, are deployed on the tablet, obviating the need for any paper case report forms that must be manually entered by a study coordinator into the trials' electronic data capture (EDC) system. Instead, a messaging service between the tablet's centralized and cloud-based database transmits the information directly to the EDC system, (which in VERVE, is hosted by another vendor). Additional aspects to the electronic capture of the patient's signature also enables efficient capture of the patient's permission for two additional important elements of the study authorization, a HIPAA authorization and a medical record release form, used for the purposes described below. As one final feature of the study that was facilitated by use of the tablet-based system, if a patient presents within the 6 weeks following receipt of the zoster vaccine (or placebo) with a rash that might be varicella infection (or less likely, shingles reactivation), the camera on the iPad is used to take one or more photographs of the rash. A graphical dermatome map is also captured via the iPad system to illustrate the distribution of the rash and evaluate whether it confirms to

one or multiple dermatomes. These images are automatically incorporated into the study database and can be sent immediately to the trial's safety monitor and as needed, to the Data Safety and Monitoring Board (DSMB) or an event adjudication committee.

Data Linkages

The VERVE trial requires only two in-person study visits, greatly increasing efficiency, reducing cost, and decreasing both study site and participant burden. However, one might wish to know about the longer-term effectiveness of the vaccine to prevent varicella zoster virus reactivation (i.e. shingles). Therefore, efficient collection of long-term outcome data on the incidence of shingles in the vaccinated vs. unvaccinated patients is important. Moreover, it might be useful to obtain similar information for patients who were eligible for the trial but elected not to participate, and on those who were never approached to participate. These two additional comparator groups can provide additional information regarding the generalizability of the trial's intervention to the population to which the intervention will ultimately be applied. Through data linkage, this information and other information from different sources (e.g. administrative claims) can be integrated and a richer and more complete data set can be generated.

Sources of Linkable data For Trial Outcome Ascertainment

A variety of data sources are available to assist in generation of real-world evidence, and these can be potentially linked to one another. Examples of potential sources of data that can inform research (and clinical care) are shown in Figure 4. While each of these has both strengths and weakness as a data source, the strengths of one data source can potentially be used to fill in the gaps in the others and supplement the data collection efforts of the trial. For safety outcomes or clinical outcomes, for example, administrative claims data from commercial health plans or governmental sources (e.g. the U.S. Medicare program) can provide a highly efficient method for case ascertainment.

Certain types of events have been shown to be accurately ascertained in administrative claims data, based on algorithms validated against medical records as the gold standard. Examples of these outcomes and their validity as ascertained in high quality validation studies are shown in Table 1. The potential benefit to this approach is that this linkage is 'passive', requiring no additional study visits, reducing participant and site burden and trial costs. If the positive predictive value (PPV) of these events are high enough (e.g. 85% or better), this approach for case ascertainment might be used by itself. In contrast, if it is not that high, then the linked claims data might be used for initial case finding, and then additional review (via the EMR or other forms of medical record retrieval), with adjudication as needed, can be conducted. This approach to use health plan claims data to augment case ascertainment has been demonstrated successfully in linkages with traditional registries (25–27) and observational studies that satisfy FDA regulatory requirements (28). This approach can also be used to support long-term safety and clinical outcome assessment in pragmatic trials. A linkage with administrative health plan data was coupled to the Women's Health Initiative (WHI) trial and was found to be able to yield essentially the same trial result as the much costlier and labor intensive cardiovascular case ascertainment and

clinical adjudication that was principally used for the WHI trial (29, 30). This capacity was also built into the VERVE trial to facilitate long-term case ascertainment for herpes zoster events, as described below.

Data Linkage in the VERVE trial

In addition to facilitating obtaining informed consent from participants, the tablet-based system described above that was implemented in the VERVE trial also digitally captured identifiable information that could be used to link to health plan data (e.g. social security number [SSN]) such as Medicare, commercial claims data, or EMR data sources. Because the validity of case identification for herpes zoster reactivation is high (PPV 85–100%) in these data sources (31–34), explicit in-person study visits to capture these events may not be required, reducing costs and minimizing participant burden. If administrative claims data is used for case ascertainment (e.g. ICD9/ICD10 code for herpes zoster, plus proximate receipt of an anti-viral medication commonly used to treat varicella) yet additional clinical information is desired, information from medical record can be obtained.

To this end, the patient's signature collected digitally in VERVE also satisfied the requirements for HIPAA authorization and was used to populate a study-specific medical record release form. The purpose of the medical record release form was to enable centralized safety follow-up, including medical record retrieval, for serious adverse events and for shingles. Specifically, if a serious adverse event occurred after 6 weeks, the VERVE call center (run by FORWARD, the National Databank for Rheumatic Diseases) was responsible for obtaining the relevant information including medical records from a provider or a hospital. If medical records were required, the electronically signed medical record release form therefore would already be available to the call center to facilitate retrieval.

In addition to the capacity to identify herpes zoster cases amongst trial participants using passive claims or EMR data sources, this approach also facilitates ascertainment of herpes zoster outcome events in non-participants. This includes both those who might have been approached at the trial site but who did not participate, as well receiving care at sites not participating in the trial. Thus, helping to inform generalizability of the trial's outcomes to the broader patient population of individuals who might benefit from the intervention.

Technical Aspects to Data Linkage- the Role of Unique Identifiers and Other Probabilistic Approaches

Although VERVE collected trial participants to facilitate data linkage with national claims and EMR data, there are many circumstances where participants will be unwilling to provide this information. Other sensitive identifiers (e.g. health insurance number) may also provoke similar reluctance for sharing. For that reason, a variety of less sensitive and non-unique identifiers can be collected to facilitate data linkage. For example, the combination of name, sex, date of birth, healthcare provider (e.g. U.S. physician's national provider ID (35) or hospital identifier), and one or more event dates (e.g. date of physician office visit, hospital admission, or device implantation) has been shown to effect data linkage between claims data and a research registry with high accuracy (36). Linkage can be done with either deterministic (i.e. rule-based) or probabilistic methods. Patient names likely need to be

normalized to remove spaces and truncate them (e.g. using the first 2 digits), since name misspellings, colloquialisms (e.g. 'Rob' rather than 'Robert'), suffixes (e.g. Jr., Sr.), and spacing variability (e.g. 'van der Heide' vs. 'vanDerHeide') will reduce the sensitivity of matching based on full name. In one study (37), a composite identifier based on truncated name and date of birth was shown to have good performance characteristics (sensitivity 97%; specificity approaching 100%) and outperformed linkage based on SSN. It also had much better sensitivity than linkage on full name and date of birth (sensitivity 87%). Evaluating this concept in the context of a university health system providing care to approximately 3 million patients, and adding in information regarding physician National Provider Identifier (NPI) as a potential identifier, the specificity of a composite identifier consisting of the first two digits of patient first and last name, middle initial, date of birth, sex, and provider NPI yielded specificity of greater than 99.8% (38).

While health systems and medical providers are generally willing to share identifiable information if patients have explicitly consented to such disclosure, there are other research-related use cases (including pragmatic trials) where this may not be feasible to obtain or was considered only after the fact and where consent is not possible to obtain post-hoc. Therefore, health systems and other covered entities will be appropriately reluctant or unable to share any personal identifying information (PII). A third party 'honest broker' can sometimes be engaged to serve in role of a trusted intermediary whose sole function is to identify and link patients between two or more data systems based on PII, but who are given no other health information (i.e. personal health information, or PHI) (39). As an alternative, cryptographic hashing approaches can be used to link patients. A hash function takes an input string (e.g. a set of patient identifiers) that applies a mathematical algorithm to create a string of a fixed size. They are typically one-way, which means that they cannot be used to reconstruct the original input string. The properties of an optimal hash function include the expectation that it is deterministic (i.e. the same inputs always results in the same hash string output), it is unique (different input strings should yield unique hash output strings), and that the hashed output is uncorrelated with its input string, so that small changes in the input string will generally result in major changes in the output string). One must first decide on a vector of patient identifiers that will be used to create the hash. This can then be 'salted' by concatenating additional a customized phrase to the input string to foil a dictionary attack. If dates must be stored (which are allowed under the rules applicable to a Limited Data Set), they can be encrypted, but then transformed to a fully de-identified dataset by setting the first observation to be an arbitrary day 0, and representing every subsequent date as the number of days offset from day 0. An example from actual implementation of a tablet-based (iPad) system used to collect patient reported outcome (PRO) data in a rheumatology clinic setting (40, 41) is shown in Table 2. The hashed string is generated by both parties, can be exchanged (since it can no longer be linked back to an individual), and then linked via deterministic methods to match patients uniquely between the data sources.

As a final concept and a limitation to the cryptographic hashing approach, the process is essentially a 'black box' and lacks verifiability of the accuracy of the approach, with opaqueness around both its sensitivity and specificity for any given linkage use case. For that reason, additional privacy conserving methods to enable and verify high quality linkage, enable minimal PII and PHI disclosure, and to monitor for drifts in data linkage have been

developed (42). This approach can take various forms, but for example, it might highlight data only where differences in the PII between potentially linked patients exist but not disclose the actual differences in the PII. This enables partial disclosure and only for data elements that are minimally necessary. The process also can support manual review for key subgroups of potentially linked matched pairs where greater uncertainty exists, increasing confidence both in the process and in its result.

Conclusion

A variety of tools and approaches are needed to support efficient generation of high quality, real world evidence to support research and promote high quality clinical care. Pragmatic trials, digital data capture including electronic consent, and linkages between data sources can facilitate these efficiencies to support the next generation of research both for rheumatology and more broadly. In addition, through platforms like SMART IRB, and with updated NIH guidance including recent changes to the Common Rule, the IRB process is becoming more efficient, creating cost and time saving for those interested in efficient real-world evidence generation. Innovations such as these in clinical trial design have the potential to allow larger, more generalizable, and more cost-effective studies that ultimately better guide real world care.

Acknowledgements:

This work was supported by NIAMS UM1 AR065705, NIAMS 1R21AR062300, 1U34AR062891, and PCORI PRN-1306-04811A02

References

1. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med*. 2016;375(23):2293–7. [PubMed: 27959688]
2. Greenberg JD, Kishimoto M, Strand V, Cohen SB, Olenginski TP, Harrington T, et al. Tumor necrosis factor antagonist responsiveness in a United States rheumatoid arthritis cohort. *Am J Med*. 2008;121(6):532–8. [PubMed: 18501236]
3. Vashisht P, Sayles H, Cannella AC, Mikuls TR, Michaud K. Generalizability of Patients With Rheumatoid Arthritis in Biologic Agent Clinical Trials. *Arthritis Care Res (Hoboken)*. 2016;68(10):1478–88. [PubMed: 26866293]
4. Harrold LR, Reed GW, John A, Barr CJ, Soe K, Magner R, et al. Cluster-Randomized Trial of a Behavioral Intervention to Incorporate a Treat-to-Target Approach to Care of US Patients With Rheumatoid Arthritis. *Arthritis Care Res (Hoboken)*. 2018;70(3):379–87. [PubMed: 28544704]
5. Congress US. 21st Century Cures Act - Public Law 114–255. 114th United States Congress; 2016.
6. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*. 2008;337:a2390. [PubMed: 19001484]
7. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147. [PubMed: 25956159]
8. Reynolds RF, Lem JA, Gatto NM, Eng SM. Is the large simple trial design used for comparative, post-approval safety research? A review of a clinical trials registry and the published literature. *Drug Saf*. 2011;34(10):799–820. [PubMed: 21879776]
9. Strom BL, Faich GA, Reynolds RF, Eng SM, D'Agostino RB, Ruskin JN, et al. The Ziprasidone Observational Study of Cardiac Outcomes (ZODIAC): design and baseline subject characteristics. *J Clin Psychiatry*. 2008;69(1):114–21. [PubMed: 18312045]

10. Officers A, Coordinators for the ACRGTA, Lipid-Lowering Treatment to Prevent Heart Attack T. Major outcomes in moderately hypercholesterolemic, hypertensive patients randomized to pravastatin vs usual care: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT-LLT). *JAMA*. 2002;288(23):2998–3007. [PubMed: 12479764]
11. Mikuls TR, Cheetham TC, Levy GD, Rashid N, Kerimian A, Low KJ, et al. Adherence and Outcomes with Urate-Lowering Therapy: A Site-Randomized Trial. *Am J Med*. 2018.
12. [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/NCT02538341) 2015; Pages <https://clinicaltrials.gov/ct2/show/NCT02538341> on 1 23, 2019.
13. i2b2.org 2019;Pages. Accessed at A National Center for Biomedical Computing at <https://www.i2b2.org/resrcs/on> January 24, 2019.
14. [TransmartFoundation.org](https://transmartfoundation.org/) 2019; Pages <https://transmartfoundation.org/> on 1 24, 2019.
15. <b/>(NIH) NIOH. Final NIH Policy on the Use of a Single Institutional Review Board for Multi-Site Research. In: (NIH) NIOH, ed. NOT-OD-16-094; 2016.
16. Messner DA, Moloney R, Warriner AH, Wright NC, Foster PJ, Saag KG. Understanding practice-based research participation: The differing motivations of engaged vs. non-engaged clinicians in pragmatic clinical trials. *Contemp Clin Trials Commun*. 2016;4:136–40. [PubMed: 29736476]
17. Williams CM, Maher CG, Hancock MJ, McAuley JH, Lin CW, Latimer J. Recruitment rate for a clinical trial was associated with particular operational procedures and clinician characteristics. *J Clin Epidemiol*. 2014;67(2):169–75. [PubMed: 24183609]
18. Eapen ZJ, Lauer MS, Temple RJ. The imperative of overcoming barriers to the conduct of large, simple trials. *JAMA*. 2014;311(14):1397–8. [PubMed: 24715072]
19. Warner ET, Glasgow RE, Emmons KM, Bennett GG, Askew S, Rosner B, et al. Recruitment and retention of participants in a pragmatic randomized intervention trial at three community health clinics: results and lessons learned. *BMC Public Health*. 2013;13:192. [PubMed: 23496916]
20. Wright NC, Foster PJ, Mudano AS, Melnick JA, Lewiecki EM, Shergy WJ, et al. Assessing the feasibility of the Effectiveness of Discontinuing Bisphosphonates trial: a pilot study. *Osteoporos Int*. 2017;28(8):2495–503. [PubMed: 28540506]
21. California So. Protection of Human Subjects in Medical Experimentation Act In: California So, ed. Health and Safety Code: California Legislature; 1978.
22. Justice SoCDo 2019;Pages. Accessed at Office of the Attorney General at <https://oag.ca.gov/research/consent> on January 23, 2019 2019.
23. Warriner AH, Foster PJ, Mudano A, Wright NC, Melton ME, Sattui SE, et al. A pragmatic randomized trial comparing tablet computer informed consent to traditional paper-based methods for an osteoporosis study. *Contemp Clin Trials Commun*. 2016;3:32–8. [PubMed: 29736454]
24. Services UDoHH. Use of Electronic Informed Consent: Questions and Answers In: Services UDoHH, ed: US Department of Health & Human Services; 2016.
25. Curtis JR, Baddley JW, Yang S, Patkar N, Chen L, Delzell E, et al. Derivation and preliminary validation of an administrative claims-based algorithm for the effectiveness of medications for rheumatoid arthritis. *Arthritis Res Ther*. 2011;13(5):R155. [PubMed: 21933396]
26. Kumamaru H, Judd SE, Curtis JR, Ramachandran R, Hardy NC, Rhodes JD, et al. Validity of claims-based stroke algorithms in contemporary Medicare data: reasons for geographic and racial differences in stroke (REGARDS) study linked with medicare claims. *Circ Cardiovasc Qual Outcomes*. 2014;7(4):611–9. [PubMed: 24963021]
27. Schousboe JT, Paudel ML, Taylor BC, Virnig BA, Cauley JA, Curtis JR, et al. Magnitude and consequences of misclassification of incident hip fractures in large cohort studies: the Study of Osteoporotic Fractures and Medicare claims data. *Osteoporos Int*. 2013;24(3):801–10. [PubMed: 23208073]
28. Xue F, Ma H, Stehman-Breen C, Haller C, Katz L, Wagman RB, et al. Design and methods of a postmarketing pharmacoepidemiology study assessing long-term safety of Prolia(R) (denosumab) for the treatment of postmenopausal osteoporosis. *Pharmacoepidemiol Drug Saf*. 2013;22(10):1107–14. [PubMed: 23857864]
29. Lakshminarayan K, Larson JC, Virnig B, Fuller C, Allen NB, Limacher M, et al. Comparison of Medicare claims versus physician adjudication for identifying stroke outcomes in the Women's Health Initiative. *Stroke*. 2014;45(3):815–21. [PubMed: 24525955]

30. Hlatky MA, Ray RM, Burwen DR, Margolis KL, Johnson KC, Kucharska-Newton A, et al. Use of Medicare data to identify coronary heart disease outcomes in the Women's Health Initiative. *Circ Cardiovasc Qual Outcomes*. 2014;7(1):157–62. [PubMed: 24399330]
31. Yawn BP, Gilden D. The global epidemiology of herpes zoster. *Neurology*. 2013;81(10):928–30. [PubMed: 23999562]
32. Baxter R, Bartlett J, Fireman B, Marks M, Hansen J, Lewis E, et al. Long-Term Effectiveness of the Live Zoster Vaccine in Preventing Shingles: A Cohort Study. *Am J Epidemiol*. 2018;187(1):161–9. [PubMed: 29309521]
33. Galil K, Choo PW, Donahue JG, Platt R. The sequelae of herpes zoster. *Archives of Internal Medicine*. 1997;157(11):1209–13. [PubMed: 9183232]
34. Donahue JG, Choo PW, Manson JE, Platt R. The incidence of herpes zoster. *Arch Intern Med*. 1995;155(15):1605–9. [PubMed: 7618983]
35. CMS.gov;Pages <https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvdentStand/DataDissemination.html> on January 23, 2019.
36. Curtis JR, Chen L, Bharat A, Delzell E, Greenberg JD, Harrold L, et al. Linkage of a de-identified United States rheumatoid arthritis registry with administrative data to facilitate comparative effectiveness research. *Arthritis Care Res (Hoboken)*. 2014;66(12):1790–8. [PubMed: 24905637]
37. Weber SC, Lowe H, Das A, Ferris T. A simple heuristic for blindfolded record linkage. *J Am Med Inform Assoc*. 2012;19(e1):e157–61. [PubMed: 22298567]
38. Mathur PSL, Bharat A, Wang S, Heckler C, Lele O, Embi PJ, Curtis JR. High Level Architecture and Evaluation of Patient Linkages for READY - An Electronic Measurement Tool for Rheumatoid Arthritis. *AMIA Annual Symposium Proceedings San Francisco, CA*; 2015.
39. Choi HJ, Lee MJ, Choi CM, Lee J, Shin SY, Lyu Y, et al. Establishing the role of honest broker: bridging the gap between protecting personal health data and clinical research efficiency. *PeerJ*. 2015;3:e1506. [PubMed: 26713253]
40. Yen PY, Lara B, Lopetegui M, Bharat A, Ardoin S, Johnson B, et al. Usability and Workflow Evaluation of “RhEumAtic Disease activitY” (READY). A Mobile Application for Rheumatology Patients and Providers. *Appl Clin Inform*. 2016;7(4):1007–24. [PubMed: 27803949]
41. Mathur P, Embi PJ, Curtis JR;Pages. Accessed at iTunes at <https://itunes.apple.com/us/app/rheumatic-diseaseactivity/id657411562?mt=8> on January 24, 2019.
42. Kum HC, Krishnamurthy A, Machanavajjhala A, Reiter MK, Ahalt S. Privacy preserving interactive record linkage (PIRL). *J Am Med Inform Assoc*. 2014;21(2):212–20. [PubMed: 24201028]
43. West SL, Storm BL, Poole C. Validity of Pharmacoepidemiologic Drug and Diagnosis Data In: Strom BL, ed. *Pharmacoepidemiology*. 4th ed. West Sussex, England: Wiley; 2005:743–55.
44. Kim SC, Solomon DH, Rogers JR, Gale S, Kleerman M, Sarsour K, et al. Cardiovascular Safety of Tocilizumab Versus Tumor Necrosis Factor Inhibitors in Patients With Rheumatoid Arthritis: A Multi-Database Cohort Study. *Arthritis Rheumatol*. 2017;69(6):1154–64. [PubMed: 28245350]
45. Xie F, Yun H, Levitan EB, Muntner P, Curtis JR. Tocilizumab and the risk for cardiovascular disease: a direct comparison among biologic disease-modifying antirheumatic drugs for rheumatoid arthritis patients. *Arthritis Care Res (Hoboken)*. 2018.
46. Curtis JR, Xie F, Yang S, Bernatsky S, Chen L, Yun H, et al. Herpes Zoster in Tofacitinib: Risk is Further Increased with Glucocorticoids but not Methotrexate. *Arthritis Care Res (Hoboken)*. 2018.
47. Cheetham TC, Marcy SM, Tseng HF, Sy LS, Liu IL, Bixler F, et al. Risk of Herpes Zoster and Disseminated Varicella Zoster in Patients Taking Immunosuppressant Drugs at the Time of Zoster Vaccination. *Mayo Clin Proc*. 2015;90(7):865–73. [PubMed: 26051268]
48. Le HV, Poole C, Brookhart MA, Schoenbach VJ, Beach KJ, Layton JB, et al. Effects of aggregation of drug and diagnostic codes on the performance of the high-dimensional propensity score algorithm: an empirical example. *BMC Med Res Methodol*. 2013;13:142. [PubMed: 24245772]
49. Curtis JR, Sarsour K, Napalkov P, Costa LA, Schulman KL. Incidence and complications of interstitial lung disease in users of tocilizumab, rituximab, abatacept and anti-tumor necrosis factor alpha agents, a retrospective cohort study. *Arthritis Res Ther*. 2015;17:319. [PubMed: 26555431]

50. Xie F, Yun H, Bernatsky S, Curtis JR. Brief Report: Risk of Gastrointestinal Perforation Among Rheumatoid Arthritis Patients Receiving Tofacitinib, Tocilizumab, or Other Biologic Treatments. *Arthritis Rheumatol*. 2016;68(11):2612–7. [PubMed: 27213279]
51. Balasubramanian A, Zhang J, Chen L, Wenkert D, Daigle SG, Grauer A, et al. Risk of subsequent fracture after prior fracture among older women. *Osteoporos Int*. 2019;30(1):79–92. [PubMed: 30456571]
52. Setoguchi S, Solomon DH, Weinblatt ME, Katz JN, Avorn J, Glynn RJ, et al. Tumor necrosis factor alpha antagonist use and cancer in patients with rheumatoid arthritis. *Arthritis Rheum*. 2006;54(9):2757–64. [PubMed: 16947774]
53. George MD, Baker JF, Hsu JY, Wu Q, Xie F, Chen L, et al. Perioperative Timing of Infliximab and the Risk of Serious Infection After Elective Hip and Knee Arthroplasty. *Arthritis Care Res (Hoboken)*. 2017;69(12):1845–54. [PubMed: 28129484]
54. Curtis JR, Chen L, Greenberg JD, Harrold L, Kilgore ML, Kremer JM, et al. The clinical status and economic savings associated with remission among patients with rheumatoid arthritis: leveraging linked registry and claims data for synergistic insights. *Pharmacoepidemiol Drug Saf*. 2017;26(3):310–9. [PubMed: 28028867]
55. England BR, Sayles H, Michaud K, Thiele GM, Poole JA, Caplan L, et al. Chronic lung disease in U.S. Veterans with rheumatoid arthritis and the impact on survival. *Clin Rheumatol*. 2018;37(11):2907–15. [PubMed: 30280369]

Key Points:

1. Efficient generation of real-world evidence demands new tools to answer questions of high importance to patients, clinicians, researchers, policymakers, and other stakeholders.
2. Pragmatic trials that study well-defined outcomes in highly generalizable patient populations can provide direct evidence about risks and benefits of medical interventions.
3. Technology-based tools (e.g. apps running on mobile devices) that facilitate electronic consent for research participation can effectively enable efficient screening and recruitment for pragmatic trials and real-world studies.
4. New methods for linkages between disparate data sources can enrich the types of information available for research purposes and can accommodate a wide spectrum of constraints around the nature and extent of the identifiable information that can be shared.

Synopsis

Real-world evidence requires use of new tools and methods to support efficient evidence generation. Among those tools are pragmatic trials, utilization of central/single institutional review board and electronic consent, and data linkages between diverse types of data sources (e.g. a trial or registry to administrative claims or electronic medical record data). This manuscript reviews these topics in the context of describing several exemplar use cases specific to rheumatology and provides perspective regarding both the promise and potential pitfalls in using these tools and approaches.

Figure 1:
PRECIS-2 Wheel Illustrating the Continuum between Pragmatic and Explanatory Trials
Note: Trials that are more pragmatic generate a wheel with its circumference closer to the rim; those that are more explanatory are closer to the hub
From Loudon K, Treweek S, Sullivan F, Donnan P, et al. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147; with permission

Figure 2:
Example of iPad Tablet-Based Study Video Describing the VERVE Pragmatic Trial

Knowledge Review

Menu

Question 1 of 6

The main reason this study is being conducted is to...

View the part of the Informed Consent that this question references and answer again.

HINT

☐

determine the safety and effectiveness of the zoster vaccine in patients treated with biologic medications who are at least 50 years old

☒

find a vaccine with no side effects.

☐

help pay for shingles treatments

☐

compare different kinds of shingles medications

Figure 3:
Assessing Patients' Understanding of the VERVE Pragmatic Trial Deployed on a Tablet

Figure 4:
Potential Data Sources to Generate Real World Evidence

Table 1:

Clinical Events and Outcomes that Can Be Ascertained with High Validity* using Administrative Data
Data from West SL, Storm BL, Poole C. Validity of Pharmacoepidemiologic Drug and Diagnosis Data. In: Strom BL, ed. Pharmacoepidemiology. 4th ed. West Sussex, England: Wiley; 2005:743–55.

Event Type	Selected Examples in Rheumatology
Myocardial infarction	(44, 45)
Stroke	(44, 45)
Cardiovascular death	(45)
Herpes zoster	(46, 47)
Gastrointestinal bleeding	(48)
Interstitial lung disease	(48, 49)
Gastrointestinal perforation	(50)
Vertebral and Non-vertebral fracture	(51)
Malignancy	(52)
Medical procedures (e.g. joint replacement)	(53)
Costs	(54)
Death	(54, 55)

* based on the existence of one or more high quality validation studies comparing an administrative data-based algorithm to the gold standard of medical record review with adjudication

Examples of Hashing, Salting and Encryption to Facilitate Data Linkage, as used in a rheumatology-specific, iPad-based app (READY) (See ref 41) to collect patient reported outcome data in rheumatology clinics

Table 2:

Process	First Name	Last Name	Sex	DOB	Hashed Identifier (One Way Hash)	Visit Date	Encrypted	Observation Period	Day Offset *	PRO Score (e.g. Fatigue)
Data input into READY app (Figure 2)	Iron	Man	M	1/1/1960		3/1/2018	-	-	-	55
Stored by READY within AWS	-	-	-	-	5e6376106b147004775cfe13a62ead21ee3e66ba1579e811778813ec09fffc6e	vE12qZ9Ar1zPIXw 731WGnBo0LLPg YyULeo8oCZDfS=	Yes	-	-	55
Available within AWS following decryption, and prior to de-identification	-	-	-	-	5e6376106b147004775cfe13a62ead21ee3e66ba1579e811778813ec09fffc6e	3/1/2018	No	-	-	55
De-identified Dataset	-	-	-	-	5e6376106b147004775cfe13a62ead21ee3e66ba1579e811778813ec09fffc6e	-	-	2018Q1 *	0	55

AWS = Amazon Web Services; PRO = patient reported outcome

“.” Signifies Null/Blank

Note: No other personal identifiers are stored in AWS. The above example is exhaustive with respect to identifiers that relate to individual patients.

* The visit date is converted into an offset based on the number of days since the date of each individual's first observation in the database, which is arbitrarily set as day 0.