RESEARCH PROTOCOL:

Comparative Risk of Infection in Rheumatic Disease Patients Initiating Immunosuppressive Therapy

Table of Contents

1. List of Abbreviations

RD	rheumatic disease
VZV	varicella-zoster virus
PML	progressive multifocal leukoencephalopathy
PJP	pneumocystis pneumonia (PJP, formerly 'PCP')
MMF	mycophenolate (mycophenolic acid or mycophenolate mofetil)
RTX	rituximab
IVIG	intravenous immunoglobulin
JAKi	JAK inhibitor
AZA	azathioprine
MTX	methotrexate
CYC	cyclophosphamide
EHDEN	European Health Data and Evidence Network
OMOP	Observational Medical Outcomes Partnership
OHDSI	Observational Health Data Science and Informatics
SNOMED	Systematized Nomenclature of Medicine

2. Responsible Parties

2.1. Investigators and Authors

Investigator/Author	Institution/Affiliation
Christopher A Mecoli, MD	Johns Hopkins University School of Medicine, Baltimore, MD, USA
Will Kelly, BS	Johns Hopkins University School of Medicine, Baltimore, MD, USA
Benjamin Martin, PhD	Johns Hopkins University School of Medicine, Baltimore, MD, USA
Erik Westlund, PhD	Johns Hopkins University School of Medicine, Baltimore, MD, USA
Christelle Xiong, MS	Johns Hopkins University School of Medicine, Baltimore, MD, USA
Gowtham Rao, MD	Jansen (Johnson & Johnson)
Joel Swerdel, PhD	Jansen (Johnson & Johnson)
Elizabeth Park MD	Columbia University
Thomas Falconer MS	Columbia University
Michael George, MD MSCE	University of Pennsylvania
Brian Toy, MD	University of Southern California

Commented [CM1]:

Commented [CM2R1]: Potential additional collaborat

UTSouthwestern
Stanford
University of Minnesota
Penn State
Oxford
Manchester

Authorship will also include those who meaningfully contribute to study design, analysis and interpretation of results and subsequently contribute to the drafting of the work for publication, approving the final version of the study. Full guidance related to how to qualify for meaningful contribution can be found on the OHDSI website: https://www.ohdsi.org/wp-content/uploads/2021/07/OHDSI-Authorship-Guidance.pdf. The Responsible Parties involved in this

protocol take accountability for the overarching protocol, package development, providing assistance to sites running the analysis and ensuring site-specific governance is adhered to in all publications generated from this protocol.

2.2 Sponsor

This study will be undertaken by Observational Health Data Science and Informatics (OHDSI), an open collaboration.[1]

Participating data owners will be responsible for self-reporting any grants funding the conversion and maintenance of their OMOP CDM. Disclosures will be reported in accordance with publication policies of journals papers are submitted to. No other disclosures reported.

3. Abstract

In this study we will determine the comparative risk of infectious complications in rheumatic disease (RD) patients initiating immunosuppressive therapy through the OHDSI Community.

4. Amendments and Updates

Number	Date	Section of study protocol	Amendment or update	Reason

5. Rationale and Background

Rheumatic diseases (RDs)—including inflammatory myositis, systemic sclerosis, and systemic lupus erythematosus—are a diverse group of immune-mediated conditions that often affect multiple organ systems and are associated with significant morbidity and mortality. Due to their relative rarity, clinical research studies on RDs frequently lack sufficient power to assess comparative effectiveness or long-term outcomes, especially when conducted at single academic centers. To overcome these limitations, leveraging real-world data sources—such as electronic health records (EHRs), insurance claims, and national healthcare system registries—offers a promising approach. These sources provide large, longitudinal datasets that enable the study of RDs at scale.

A major clinical challenge in RD management is the high burden of infectious complications. These arise from both the underlying immune dysregulation inherent to RDs and the immunosuppressive therapies used to treat them. Although individually uncommon, these infections collectively represent a substantial and often underrecognized contributor to morbidity and mortality, with some associated with fatality rates exceeding 50%. Despite their clinical significance, our current ability to predict these events and stratify patient risk remains limited.

The following are examples of rare but severe infectious complications that warrant further investigation:

1. Infections requiring hospitalization: Patients with rheumatic diseases (RDs) face a significantly elevated risk of serious infections that require hospitalization, due to a combination of disease-related immune dysfunction and immunosuppressive treatments. Both the

underlying autoimmune process and medications such as corticosteroids, biologics, and cytotoxic agents impair the immune system's ability to respond effectively to pathogens. As a result, patients are more susceptible to opportunistic infections, reactivation of latent infections (e.g. herpes zoster), and severe outcomes from common community-acquired pathogens.

The burden of these complications extends beyond individual patient outcomes. Recurrent or severe infections can lead to treatment interruptions, suboptimal disease control, and increased healthcare utilization. Additionally, infectious complications often influence clinician prescribing behavior, leading to more conservative use of effective immunosuppressive regimens in high-risk patients.

2. Varicella Zoster: One out of every three Americans will develop shingles during their lifetime, and of that group, 1 in 10 patients will develop post-herpetic neuralgia (PHN) that can last months or years. Patients with RDs are at higher risk for developing VZV compared with the general population. With an increasing use of higher dose and combination immunosuppressant therapy, VZV continues to be challenge to physicians treating patients.

3. Progressive multifocal leukoencephalopathy (PML): Both mycophenolate and rituximab are often used to treat RDs, and both medications have been issued black box warnings from the FDA for the development of PML. Historically, most data on PML have focused on HIV and oncology patient populations, the latter often extrapolated to rheumatic diseases. In oncology patients, the estimated incidence was reported to range from 2-10/100,000 patient years [2]. PML in oncology patients carries a median survival of 100 days, and for those that survive long term there is often residual cognitive deficits [3]. As most of our data on PML originates from single center studies, the ability to examine specific medication regimens and disease phenotypes has not been possible to date.

4. Pneumocystis jirovecii pneumonia (PJP): PJP is an opportunistic infection with high mortality among patients with underlying rheumatologic conditions, ranging from 33 to 60% depending on the cohort examined. Despite these mortality rates, up to 30% of rheumatologists and other healthcare providers do not routinely administer prophylaxis [4]. While corticosteroids are now a well-established risk factor for the development of PJP, other risk factors (or protective factors) are not well established.

6. Objectives

The primary objective of this study is to determine the risk of infectious outcomes in patients across several autoimmune conditions on both monotherapy and combination therapy.

7. Methods

7.1 Data Sources

This study is a multinational cohort study. We intend to solicit participation from a variety of healthcare settings in multiple geographies (**Table 1**). Should more data partners wish to participate,

this analysis could extend to any additional databases that are formatted to the Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM). We welcome those with data mapped to the OMOP-CDM to join us in executing this study if they are willing to participate.

The study will be conducted using data from real world data sources that have been mapped to the OMOP-CDM in collaboration with the Observational Health Data Sciences and Informatics (OHDSI) and European Health Data and Evidence Network (EHDEN) initiatives. The OMOP-CDM (https://github.com/OHDSI/CommonDataModel/wiki) includes a standard representation of health care experiences (such as information related to drug utilization and condition occurrence), as well as common vocabularies for coding clinical concepts, and enables consistent application of analyses across multiple disparate data sources.

Table 1. Participating Data sources mapped to the OMOP CDM

Data source	Source population	Sample size	Data type	Longitudinal history
Columbia University Irving Medical Center	Patients of the Columbia University Irving Medical Center (New York City, USA)	≈ 6 million	General practice electronic health records, outpatient specialist electronic health records, inpatient hospital electronic health records, hospital billing/summary	1989 (1978 for diagnoses)
Johns Hopkins	EHR records	≈2.4 million	General practice electronic health records, outpatient specialist electronic health records, inpatient hospital electronic health records, hospital billing/summary	2016

7.2 Study design and questions to be answered

The study will be a retrospective observational cohort study based on routinely collected health care data which has been mapped to the OMOP-CDM. We will utilize the Characterization and PLE packages via Strategus/HADES to execute this study to answer the following questions:

7.2.1.
Among patients with prevalent rheumatic disease X [where X = systemic lupus, systemic sclerosis, inflammatory myositis, and non-infectious uveitis] on new-start immunosuppressant Y [where Y = MMF, RTX, IVIG, or JAKi] what is the incidence rate of the infectious outcome Z [where Z = PML, PJP, VZV, and hospital-required infection]?

7.2.2.

Among patients with prevalent rheumatic disease X [where X = systemic lupus, systemic sclerosis, inflammatory myositis, and non-infectious uveitis] on prevalent immunosuppressant 1 [where 1 = MMF, MTX, AZA] **who add on** immunosuppressant 2 [additional agent] what is the incidence rate of the infectious outcome Z [where Z = PML, PJP, VZV, and hospital-required infection]?

7.3 Study Population

We have previously created several RD phenotypes (i.e. case identification algorithms) through the Observational Health Data Sciences and Informatics (OHDSI) Community. These include both sensitive and specific phenotypes for systemic sclerosis, dermatomyositis, systemic lupus erythematosus, and non-infectious uveitis. These phenotypes will be used for the current study.

7.4 Exposures

Exposures include medications (immunosuppressants and immunomodulatory agents), both as monotherapy and as combination therapy

7.5 Outcomes

- Hospitalized Infection
- Progressive multifocal leukoencephalopathy (PML)
- Pneumocystis jirovecii pneumonia (PJP)
- Shingles (VZV)

7.6 Study Design

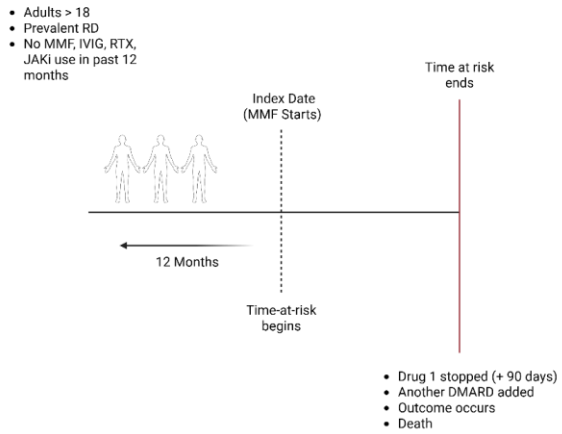
All analyses will be performed using code developed for HADES (Health Analytics Data-to-Evidence Suite), formerly known as the OHDSI Methods library.

Aim 1: To evaluate the comparative safety of mycophenolate (MMF) vs rituximab (RTX) vs. intravenous immunoglobulin (IVIG) vs JAKi (tofacitinib, baricitinib) in adult autoimmune disease (DM, SSc, SLE, uveitis).

Hypothesis: More infections will be observed in JAKi and MMF compared to RTX, with IVIG being safest.

Methods: New User Design. Identify adults ≥18 with the rheumatic disease of interest (prevalent disease) who have not received any of these four therapies in the past 12 months (but allow any other DMARDs). Index date is start of MMF, RTX, IVIG, JAKi. Follow until medication ends (with 90-day lag) or until another therapy [MMF, RTX, IVIG, JAKi, or any other DMARD] is initiated.

Expected Results: This would provide IRs for initiators of monotherapy of these agents across multiple autoimmune rheumatic diseases *(the comparative safety profiles of these drugs have never been examined).*



Rationale: For Aim 1, we selected new users of four commonly used drugs that would all be reasonable treatment options in patients with active rheumatic disease (clinical equipoise). We also chose these four drugs because little is known about their relative infectious risks between them (and across rheumatic diseases). This aim would result in quantification of incidence rates for each drug across several different autoimmune diseases, providing insight into the risk conferred by the medication as well as the risk conferred by the underlying disease.

1. Target Population

- Adults (≥18) with a diagnosis of SLE, SSc, DM, or non-infectious uveitis
- No prior use of MMF, RTX, IVIG, JAKi in the 12-month baseline period (new users)
- At least 12 months of medical history available before the index date

2. Index Date (Time Zero)

- Initiation of MMF, RTX, IVIG, or JAKi

3. Baseline Covariates

Collected during the 12-month lookback period *(Rationale: To ensure balancing the four treatment groups to ensure they are as similar as possible*)

- Demographics
- Comorbidities
- Prior immunosuppressant use
- Concurrent medications
- Healthcare utilization
- Vaccination and infection history
- Rheumatic Disease Comorbidity Index

4. Outcomes

- **Primary**: Time to infection (PCP, PML, shingles, hospitalized infection) (*Rationale: The infectious outcomes were chosen due to their significant morbidity and mortality in rheumatic diseases, as well as clinicians having the ability to modify risk [with prophylaxis, screening, or vaccination, respectively]. They were also chosen due to the ability to accurately identify these infections as they have distinct SNOWMED codes).*

5. Follow-up

Start at index date and continue until:

- Outcome occurs
- End of medication exposure
    - EHR end-of-exposure: (a) discontinue date of medication with no renewal/refill within 30 days for MMF and JAKi; (b) for infusions (RTX, IVIG): infusion date (+28 days for IVIG; +180 days for RTX), add 90 days = treatment ended; (c) OR, if not available: (c) HADES treatment pathways model: persistence window 30 days (drug start until drug end based on discontinuation or supply prescribed) with no drug renewed/refilled for 90 days
    - Claims end-of-exposure: For prescriptions (MMF, JAKi): (a) medication fill date + medication days supply → calculate end date; if > 90-day gap until next RX → treatment ended. For infusions (RTX, IVIG): infusion date (+28 days for IVIG; +180 days for RTX), add 90 days = treatment ended.
- Death
- Study period ends

6. Analysis Strategy

Use multinomial logistic regression to create a propensity score with a multi-level exposure (the 4 treatment groups) and create inverse probability treatment weights so everyone can be compared in one model. Use available covariates in 12 month look back period to derive propensity score.

- Subgroup analyses
    - Prednisone dose at index date (calculated prior month median daily dose)
    - Shingles vaccination (Zostavax/Shingrix) at index date: censor if/when vaccinated (using all available data to capture vaccine history)
    - PCP prophylaxis (TMP/SMX, atovaquone, dapsone) defined as medication fill within 90 days of index date
    - *(Rationale: We elected for our primary analysis to include these covariates at only baseline for simplicity, given the complexities introduced by their time-varying nature. For shingles vaccination, patients who were vaccinated during follow-up period were censored at date of vaccination)*

- Sensitivity analyses:
    - 1. Allow patients to enter cohort even if they have received MMF, IVIG, RTX, or JAKi in the preceding 12 months

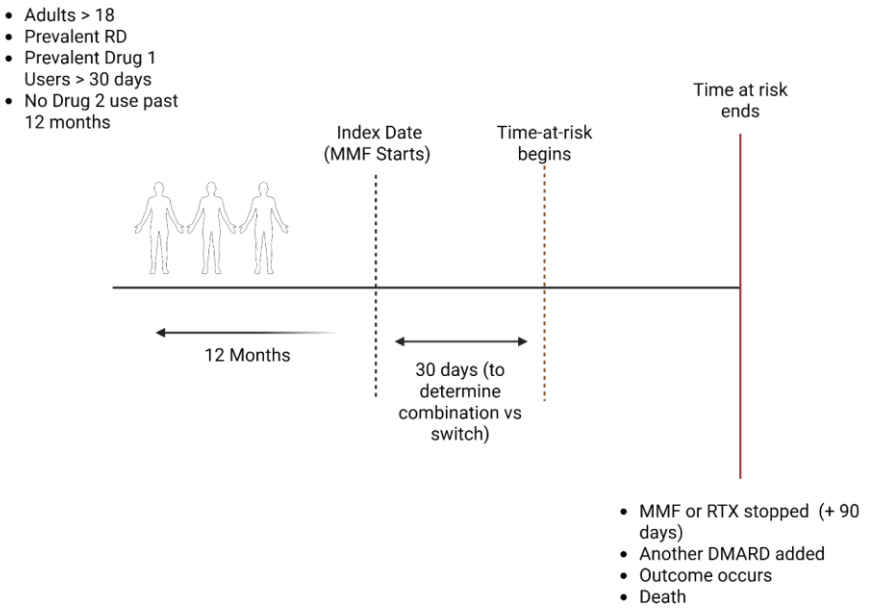
- *(Rationale: This would likely increase the number of individuals eligible for analysis, as well as mimic more real-life practice)*

- 2. Modify lag period to 0 and 120 days
- 3. Pool all RDs together
  - *(Rationale: Gain statistical power but introduce heterogeneity inherent to underlying rheumatic disease)*
- 4. Pool EHR vs claims data sources
- 5. Incorporate time-varying covariates for prednisone dose (none, low-dose, medium, high) and PCP prophylaxis

Aim 2: To evaluate the comparative safety of different combinations of immunosuppressive therapies among adults with autoimmune disease on prevalent therapy with one agent who add on a second agent.

Methods: New User/Active Comparator Design. Identify adults ≥18 with rheumatic disease of interest (prevalent disease) who are prevalent (>30 days) users of MMF with no RTX use for the past 12 months. Index date is start of RTX while continuing MMF. Time at risk continues until exposure of either MMF or RTX ends (with 90-day lag) or until another DMARD therapy is initiated.

Expected Results: This would provide HRs examining different combinations of conventional and biologic DMARD therapies used across RDs.



(Rationale: For Aim 2, we selected prevalent users of three commonly used DMARD agents (MMF, MTX, AZA) who escalated therapy with a second agent. Prevalent medication users were selected to

capture the impact of the combination (adding the second agent). This aim will inform clinicians and patients when deciding on which second agent to choose for a given rheumatic disease).

1. Target Population

- Adults (≥18) with a diagnosis of SLE, SSc, DM, non-infectious uveitis
- Prevalent use of Drug A (e.g. MMF) (does not need to be new user)
- No prescription for Drug B (e.g. RTX) within at least 12 months

2. Index Date (Time Zero)

- Date of first prescription of Drug B (RTX)

3. Baseline Covariates

Collected during the 12-month lookback period:

- Demographics
- Comorbidities
- Prior immunosuppressant use
- Concurrent medications
- Healthcare utilization
- Vaccination and infection history
- RDCI

4. Exposure Groups
With the exact same study design, look at the additional medication combinations (first drug prevalent drug, second drug initiator):

Details of defining a combination from HADES package:
combinationwindow: 30 days
minpostcombinationduration: how long each drug must continue to be counted as true combination (30 days). Sensitivity analysis of 90 days. If <30 or < 90, exclude from analysis as this would be counted as a "switch" or discontinuation of Drug 1.

Parameter	Meaning
combinationWindow	Max days between drug starts to be called a combo
minPostCombinationDuration	Min days both drugs must persist to count as a combo
Drug A stops and Drug B starts later	= **Switch**
Drug A ongoing when Drug B starts within combo window	= **Combination** (if duration threshold met)

Combinations of Interest:
-MMF + IVIG
-MMF + MTX
-MMF + AZA
-MMF + JAKi

-MTX + IVIG
-MTX + MMF
-MTX + AZA
-MTX + RTX
-MTX + JAKi

-AZA + IVIG
-AZA + MMF
-AZA + MTX
-AZA + RTX
-AZA + JAKi

4. Outcomes

- **Primary**: Time to infection (PCP, PML, shingles, hospitalized infection) (*Rationale: The infectious outcomes were chosen due to their significant morbidity and mortality in rheumatic diseases, as well as clinicians having the ability to modify risk [with prophylaxis, screening, or vaccination, respectively]. They were also chosen due to the ability to accurately identify these infections as they have distinct SNOWMED codes).*

5. Follow-up

Start at index date and continue until:

- Outcome occurs
- End of medication exposure
    - EHR end-of-exposure: (a) discontinue date of medication with no renewal/refill within 30 days for MMF and JAKi; (b) for infusions (RTX, IVIG): infusion date (+28 days for IVIG; +180 days for RTX), add 90 days = treatment ended; (c) OR, if not available: (c) HADES treatment pathways model: persistence window 30 days (drug start until drug end based on discontinuation or supply prescribed) with no drug renewed/refilled for 90 days
    - Claims end-of-exposure: For prescriptions (MMF, JAKi): (a) medication fill date + medication days supply → calculate end date; if > 90-day gap until next RX → treatment ended. For infusions (RTX, IVIG): infusion date (+28 days for IVIG; +180 days for RTX), add 90 days = treatment ended.
- Death
- Study period ends

6. Analysis Strategy

Use multinomial logistic regression to create a propensity score with a multi-level exposure (the 4 treatment groups) and create inverse probability treatment weights so everyone can be compared in one model. Use available covariates in 12 month look back period to derive propensity score.

- Subgroup analyses

- Prednisone dose at index date (calculated prior month median daily dose)
- Shingles vaccination (Zostavax/Shingrix) at index date: censor if/when vaccinated (using all available data to capture vaccine history)
- PCP prophylaxis (TMP/SMX, atovaquone, dapsone) defined as medication fill within 90 days of index date
- *(Rationale: We elected for our primary analysis to include these covariates at only baseline for simplicity, given the complexities introduced by their time-varying nature. For shingles vaccination, patients who were vaccinated during follow-up period were censored at date of vaccination)*

- Sensitivity analyses:
  - 1. Modify lag period to 0 and 120 days
  - 2. Pool all RDs together
    - *(Rationale: Gain statistical power but introduce heterogeneity inherent to underlying rheumatic disease)*
  - 3. Pool EHR vs claims data sources
  - 4. Incorporate time-varying covariates for prednisone dose (none, low-dose, medium, high) and PCP prophylaxis

7.7 Logistics of Executing a Federated Analysis

Sites will run the study analysis package locally on their data coded according to OMOP-CDM. Only aggregate results will be shared with the study coordinator. Result files will be automatically staged into a ZIP file that can be transmitted using the OhdsiSharing R Library (http://ohdsi.github.io/OhdsiSharing/) or through a site's preferred SFTP client using a site-specific key provisioned by the OHDSI Study Coordinator. Local data stewards are encouraged to review study parameters to ensure minCellCount function follows local governance. At a minimum, it is encouraged to keep this value to > 10 to avoid any potential issues with re-identification of patients.

8. Sample Size and Study Power

The study package will be designed to suppress any analyses which have less than 10 unique persons. This parameter is configurable and can be adjusted by the analyst executing the package, should their institution require a different threshold. This means that each data owner will only generate results for target-stratum-feature pairs that meet this minimum threshold.

9. Strengths and Limitations

9.1 Strengths

Leveraging the OHDSI network/the OMOP CDM offers a promising solution for studying rare rheumatic diseases by utilizing real-world data from diverse sources. This approach can improve the generalizability and reproducibility of research findings, ultimately contributing to a better understanding of treatment complications and inform clinical decision making.

9.2 Limitations

While network studies using multiple real-world databases offer the advantage of large sample sizes, they also introduce several methodological challenges—particularly related to data quality and phenotype accuracy.

First, identifying infectious complications can be challenging due to limited availability of microbiologic or diagnostic measurement data (e.g., PCR results), increasing the risk of outcome misclassification.

Another limitation stems from potential inconsistencies in Extract, Transform, and Load (ETL) processes when mapping source data to the OMOP CDM. Variability in these processes across sites can impact data completeness and the validity of derived phenotypes.

The identification of comorbidities or serious adverse events relies on diagnostic coding. In this context, the absence of a diagnostic code is assumed to represent the absence of a condition—an assumption that likely underestimates true disease prevalence, especially for under-coded or poorly captured conditions.

Medication exposure is another potential source of misclassification. Prescription or dispensing records indicate intent to treat, but do not confirm actual medication adherence. This may lead to false positives (e.g., a patient received a prescription but did not take the drug) or false negatives (e.g., a patient continued therapy beyond the recorded days' supply due to stockpiling).

Lastly, observed differences in outcomes or treatment patterns over time may reflect evolving clinical practices or changes in data capture methods rather than true differences in patient characteristics or behavior. This is especially important to consider in comparisons of medication use across different time periods or settings.

10. Protection of Human Subjects

The study uses only de-identified data. Confidentiality of patient records will be maintained at all times. Data custodians will remain in full control of executing the analysis and packaging results. There will be no transmission of patient-level data at any time during these analyses. Only aggregate statistics will be captured. Study packages will contain minimum cell count parameters to obscure any cells which fall below allowable reportable limits. All study reports will contain aggregate data only and will not identify individual patients or physicians.

11. Plans for Disseminating and Communicating Study Results

All results will be posted on a freely available and accessible website such as the OHDSI website (evidence.ohdsi.org) after completion of the study. Results are aimed for publication in a clinically focused peered reviewed scientific journal to inform future shared patient and clinician decision making. The results will also be presented at the OHDSI in-person events.

12. References

1. OHDSI. Observational Health Data Sciences and Informatics 2021 [Available from: https://www.ohdsi.org/.
2. Amend KL, Turnbull B, Foskett N, Napalkov P, Kurth T, Seeger J. Incidence of progressive multifocal leukoencephalopathy in patients without HIV. Neurology. 2010 Oct 12;75(15):1326-32. doi: 10.1212/WNL.0b013e3181f73600. PMID: 20938025.
3. Anand P, Hotan GC, Vogel A, Venna N, Mateen FJ. Progressive multifocal leukoencephalopathy: A 25-year retrospective cohort study. Neurol Neuroimmunol Neuroinflamm. 2019 Sep 25;6(6):e618. doi: 10.1212/NXI.0000000000000618. PMID: 31554669; PMCID: PMC6814409.
4. Cettomai D, Gelber AC, Christopher-Stine L. A survey of rheumatologists' practice for prescribing pneumocystis prophylaxis. J Rheumatol. 2010 Apr;37(4):792-9. doi: 10.3899/jrheum.090843. Epub 2010 Mar 1. PMID: 20194450.