Published in final edited form as:

Commun Stat Theory Methods. 2009 January 1; 38(18): 3268–3281. doi:10.1080/03610920802468681.

On Determining the Effects of Therapy on Disease Damage in Non-randomized Studies with Multiple Treatments: A study of Juvenile Myositis

Peter A. Lachenbruch,
Oregon State University, Corvallis, OR

Frederick W. Miller, and
Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Bethesda, MD

Lisa G. Rider
Environmental Autoimmunity Group, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Bethesda, MD

For the International Myositis Assessment and Clinical Studies (IMACS) Group

Abstract

This article discusses predicting damage in patients with juvenile myositis after treatment with various medications. These data were taken from medical records and not randomized. Investigators advocate using propensity scores for analysis of such non-randomized studies in order to reduce the effect of selection of treatment. Thus far, the studies have typically been comparing drug administration versus no drug after including a propensity score to compensate for potential bias in selecting patients for use of the agent. In this study, we use propensity scoring for multiple treatments given singly or in combination. We study two methods. We use a multiple logistic regression model with continuous propensity scores and a model that develops strata based on the dichotomous treatment assignment (received drug or not). We find the multiple logistic regression models predict damage better than the dichotomous model. In many cases, the propensity score also accounts for the effect of the treatment.

Keywords

propensity scoring; multiple treatments; myositis

Introduction

Myositis is a rare group of systemic autoimmune diseases characterized by chronic muscle inflammation and weakness (Rider, et al 2007). It can affect both children and adults. Recent estimates suggest that the incidence of juvenile dermatomyositis (JDM) is 3.2 cases per 1 million children (Mendez, et al, 2003). For this reason, it is difficult to conduct randomized clinical trials of the disease. Many studies have investigated the effects of therapies in small numbers of patients, most at single referral centers. It is thus important to try to combine studies and to account for the non-randomized nature of the studies.

Juvenile myositis has an onset between 4 and 15 years of age. There are two main forms: dermatomyositis that has characteristic skin rashes (JDM), and polymyositis that does not (JPM). Girls are affected about twice as often as boys, and all racial and ethnic groups are affected (Mendez, et al 2003). Juvenile myositis usually develops over a period of weeks to months. Myositis that develops in patients older than 18 years is adult myositis. It differs from juvenile myositis in some features (Rider and Miller 2000).

Common **treatment** options for myositis are listed below (table 1) (adapted from Table 12.1 in Rider, et al). The mainstay of therapy is daily oral corticosteroid therapy (prednisone). For refractory patients or patients with severe illness onset or those with risk factors for a poorer outcome, other therapies may be used, most of which are immunosuppressive therapies. Most of these agents were not studied in randomized controlled trials.

These therapies were developed from experience with the illness, but it has been difficult to conduct clinical trials because of the rarity of the disease. Often these therapies were developed by analogy with the treatment effect observed in adult myositis patients and in other autoimmune diseases or studies based on open label or retrospective observations of benefits to the patients. All JDM patients receive corticosteroids. Physicians administer the other therapies based on their assessment of the patients. Thus, simply comparing the outcomes of a particular therapy versus those patients not on that therapy may be confounded with other indicators of patient condition.

Many studies attempt to look at this issue using covariate adjustment. However, such analyses may use most of the degrees of freedom in the data set since the size of the data sets is usually small and the number of covariates can be large. A recent method, propensity scoring, was developed to generate strata of likelihood of receiving the drug (see, e.g., D'Agostino, 1998, for a nice review). In this method, a logistic regression model with the receipt of treatment as the dependent variable and the potential confounders as predictors determines an estimated probability of receiving a drug (the propensity score). We then stratify probabilities into quintiles and use these strata as a way of adjusting for the physician's decision to give the drug. This is a standard method (see e.g., D'Agostino, 1998). One may also consider using the propensity scores directly as covariates. We did not do this, as there are potential non-linearities in the prediction of outcomes from the propensity scores. Use of the strata allows some non-linearity to be present. Users generally suggest utilizing a large number of predictors to determine the strata without adjustment for the number of predictor variables. Thus, if 20 variables were used in the development of the strata, no recognition of this fact is given in the analysis (but worth stating based on recent evidence that the number of variables in the propensity score model- and whether they are selected or unselected - does not seem to matter). Other studies suggest using the propensity score to match treated and untreated patients. This can result in using some untreated patients to match multiple patients. Rubin (1997, p.762) notes that "...at least in moderate or large samples, the biasing effects of leaving out even a weakly predictive covariate dominate the efficiency gains from not using such a covariable." Brookhart et al (2006) conducted a simulation study of selection of variables for propensity scores. They noted that there are occasions in "very small studies there are situations in which it might be advantageous to exclude true confounders from a PS model. This occurs when the confounder is only weakly related to the outcome but very strongly related to the exposure... in moderate-sized studies; one would not want to exclude any covariate related to exposure..." The analysis is conditional on the propensity score. It is not clear what the effect of this is on the uncertainty of the evaluation of the drug.

Note that these are simple analyses of drug versus not drug. However, many patients may receive several drugs (either sequentially or simultaneously). We wish to have a method that would create a propensity score that would account for multiple therapies. One solution would

be to create a “drug” variable that could have (say for 3 drugs) no drug, drug A, drug B, drug C, drugs A&B, drugs A&C, drugs B&C, drugs A& B&C. With five strata for each possible drug, using multiple logistic regression could lead to $5^8 = 390625$ categories (an obviously impractical solution). One might consider only two strata per drug, perhaps breaking at the median, or at a high propensity probability cut point – say the 80th percentile of risk. This would lead to $2^8 = 256$ categories – better, but still impractical. One might consider using the propensity scores and two- and three- factor interactions as predictors. This also leads to consideration of non-linear response to the damage indices. An alternative we consider is to use the bivariate strata and consider only the two highest strata versus the other categories for each drug. Thus, one would combine the two highest strata for each drug propensity. This gives a set of dichotomies which are one if the patient is above the 60th percentile of propensity scores for each drug. We chose the two highest strata (above the 60th percentile of the propensity score) as we had observed that when this scoring was predictive, the two highest strata were the most predictive. We then formed a prediction equation using the dichotomies and their two- and three-factor products. For three drugs, we have the following (table 2) where H60 is 1 if the propensity score for hydroxychloroquine is above the 60th percentile and 0 if not (similarly for methotrexate and intravenous immune globulin):

The variable HCQ is 1 if hydroxychloroquine was given, MTX is 1 if methotrexate was given and IVIG is 1 if intravenous immune globulin was given. Our analysis would then use the model response = $\beta_0 + \beta_1 * H60 + \beta_2 * M60 + \beta_3 * I60 + \beta_4 * H60 * M60 + \beta_5 * H60 * I60 + \beta_6 * M60 * I60 + \beta_7 * H60 * M60 * I60 + \gamma_1 * HCQ + \gamma_2 * MTX + \gamma_3 * IVIG$.

This would permit adjusting for the strata, the combination of several strata, and the administration of the drugs. In particular, if the individual drugs (HC Q, MTX, and IVIG) do not have statistically significant coefficients, the strata would account for the effects of the drugs because of the relation to the choice of drug. Note that we develop this model from the individual drug vs. no drug models and use the highest stratum from each model. We did not include interactions of HCQ, MTX and IVIG because the interactions were not significant. This is probably due to relatively low frequency of administration of these drugs multiply. HCQ was administered 8 times with MTX, IVIG was administered once with MTX and IVIG was not administered with HCQ. Thus, interactions were not useful.

The total damage visual analog scale (totalVas) score is a severity of damage score developed in the Myositis Damage Index which is the sum of several visual analogue scales (Isenberg, 2003): muscle_vas; skeletal_vas; cutaneous_vas; gastro-intestinal vas (gi_vas); pulm_vas (pulmonary); cardiovascular vas (cv_vas); peripheral vascular (vasc_vas); endocrine_vas; ocular_vas; infection_vas; malignancy VAS (ca_vas). Each of these visual analog scales measures a dimension of disease. The sum of these represents a burden-of-disease. It is accepted as a measure of disease damage.

The variables we used to predict the given treatment are in table 3. We include the variable name followed by the number of cases with that variable.

We collapsed onset severity by combining the two highest categories to form a three category variable when inclusion of the fourth category of onset severity led to collinearities in estimation and loss of observations. We also considered using the treating center in our models, but they did not contribute to prediction so we did not pursue them further. There are some problems of missing values. There were unrecorded values in the medical records for onset severity, dysphagia, cardio-pulmonary involvement, and ulceration. These were not in common, so the net reduction in complete cases could be as large as 20. In some instances we considered only dermatomyositis cases, which cause a loss of 9 cases. This leads to a more homogeneous group.

Propensity scoring to determine the effect of a Single Therapy vs. any other therapy

In this model, one obtains the probability a physician gives a treatment based on the values of the covariables. This attempts to adjust for bias in treatment assignment. Then we create strata based on the percentiles of the estimated probability that the treatment is given. The first step is to conduct a logistic regression analysis to predict any use of hydroxychloroquine during the period of observation. We excluded some observations because the predictor variables in the propensity score model had missing values. The results, given below as preliminary findings, suggest a strong ability to predict the chance of hcq being given. After conducting this logistic regression, we compute the predicted values of the probabilities and determine the quintiles of “risk” of getting hcq. These quintiles form the risk strata for the analysis of the damage variable totalVas. We can then perform ANOVA or regression modeling using the strata and the variable anyhcq to determine if giving hcq at any time during the illness has an effect above the treatment selection effect. As a model check, we also ran the model including interactions of all predictor variables with the strata. We found no significant interactions (all had p-values greater than 0.4) and the main effects for the covariates were all non-significant except for the three level onset severity.

We also did this for methotrexate (mtx) and for intravenous immune globulin (ivig) to obtain information on the efficacy of each drug individually. These preliminary results are also given in Table 4 below. We give the Stata commands in the appendix with some commentary.

The results for onset severity compare each level to the onsetseverity level 1, the mildest form. We note that only five patients had severity level 4 hinting that we should not place too much faith in this result. There are few strong predictors of treatment administration, so the effects of stratifying will be small. However, the strata may affect the drug effect.

The analysis for the effect of treatment after adjusting for the chance of getting hcq is found by an analysis of covariance. We note that the model is not a good predictor (the model below has a low R^2), so even with a fairly strong ability to predict treatment choice, the strata and the treatment together do not provide much information on damage outcome.

The output of the ANOVA command (Table 5) shows the highest risk category in the strata is dropped. In this case, it is the highest risk category. Interestingly, only stratum 4 has a lower totalVas (severity of damage) score than the highest category at risk of getting the drug. Thus, the strata are important in predicting whether the patient received the drug, but are not very important in predicting totalVas.

We can also test if the strata are helpful using a likelihood ratio test. We show the commands and output in appendix 2.

We find the univariate propensity score analyses for all drugs in this way. This is a standard method. Another method is to use the propensity score to match treated and untreated subjects. This works well when the data set is large. When it is not, some patients may not have a match, which results in a reduced sample size. The number of matches is the minimum of the number of treated and the number of control patients, or some patients match multiple times. This suggests using matched sample techniques. There is debate in the literature on whether the variance should be based on the matched pairs or the unmatched variance. In addition, there may be multiple matches of control patients and treated patients. This is not necessarily a bad thing, but the investigator should be aware of this. It does complicate the analysis. We did not use the matching method.

The propensity scores for each drug provide a method of (partially) adjusting for the features that the physician uses to select a drug. However, some patients get more than one drug, sometimes jointly and sometimes sequentially (Table 6). We wish to find a way to adjust for the propensity of multiple drug administration.

The zero category for all drugs consists of the patients who had only prednisone. The zero categories for the second line are the patients who did not have HCQ, MTX, or IVIG.

The use of the strata to predict the chance of getting a drug is important since many of these analyses suggest that getting a particular drug treatment is useful, but when we include the strata, the significance often fades away. That is, the strata predict all of the information on the totalVas as well as the drug does by itself. This is likely due to physicians selecting drugs based on the patient characteristics.

Determining the effect of multiple therapies

In practice, most patients receive multiple therapies during the course of their illness. Thus, there is overlap in the anyhcq, anymtx and anyivig variables. There are several ways of determining a propensity score for this situation. One would be to consider the multivariate distribution of all three treatments: anyhcq, anymtx, anyivig. The variable treat is defined as 0 = no therapy, 1=anyhcq, 2=anymtx, 3= anyivig, 4=anyhcq and anymtx, 5 anyhcq and anyivig, 6=anymtx and anyivig, 7 anyhcq and anymtx and anyivig. This gives eight outcomes. We use a multiple logistic model, and compute propensity scores from the prediction equations. We have multiple logistic equations for each of the treatment combinations. We found that several of the prediction variables (cardio-pulmonary involvement, onset severity=4) had zero or one outcome, so they were not useful for prediction. We collapsed onset severity into 3 categories and omitted cardio-pulmonary involvement. Treatment combination 5 (hcq & ivig) had only one patient, so we omitted that category. The command to perform multinomial logistic regression (we omit the output, as it is complicated and not informative to this paper) is given in the appendix.

We use the predict command to get six propensity scores (Table 7).

We use these to remove the effects of the physician's treatment selection process. We show the preliminary analysis using the propensity scores and the treatment given (Note that treatat5 is omitted because there was only one patient with the combination hcq + ivig. Also, we specified treatat1 (anyhcq) as the comparison group in the multiple logistic regression). We do a multiple linear regression using the propensity scores for treatment and the indicators for the drugs being given (Table 8). This gives the prediction of totalVas given the propensity scores and the treatment given. The p-value for this regression is <0.0001, with an $R^2=0.50$ indicating there is a strong relationship of these variables to totalVas. We can conduct a likelihood ratio test of the utility of the three treatment variables given the propensity scores by re-estimating the model with the same command and omitting the three treatment variables (anyhcq, anymtx, anyivig). The regression model has a p-value of <0.0001. The likelihood ratio test for the deletion of the three treatment variables gives a p-value of 0.30 indicating that the treatment indicators did not provide evidence of additional information when the propensity scores were in the model.

We also performed a likelihood ratio test of the effects of the propensity scores (omit the treat1hat-treat7hat variables, i.e. the propensity scores) and find the LR χ^2 with 5 d.f. is 24.26 with a p-value of 0.0002. This strongly suggests that the propensity scores help predict damage. We conclude that the treatment selection process is informed by the covariables to the extent that the damage effects are largely accounted for by the treatment selection process (which is something the physician does without a formal test).

These do not stratify on the predicted probabilities. To do so would require us to determine the number and levels of strata for each of the six treatment combinations risk scores. With six possible treatment combinations, we could have $2^6 = 64$ strata if we dichotomized (say at the 80th percentile) or $3^6 = 729$ with three levels – this is obviously not a feasible solution.

An alternative model we consider uses propensity strata that were found in the single-treatment analyses and find a model that combined them in some manner. We considered the dichotomization of the five-level strata mentioned earlier into the highest two and remaining three strata (called anyivig34, etc. for the third and fourth strata – they were number 0 to 4, etc.). The rationale for this would be that these treatment administrations can be described by a two-factor model. There is an apparent relationship between anymtx and the other drugs. However, using logistic regression, we can see that mtx has a strong relationship with hcq and a moderate one with ivig (table 9). That is the only treatment that had a relationship with more than two drugs.

This suggests that these indicators can be modeled using the propensity scores and interactions. This heuristic is not perfect, but provides a comparison to the multiple logistic.

We consider the regressions of the totalVas on the individual strata predictors and their two-way interactions. The variables anyivig34 – anymtx34 are the main effect indicators, ivhc-hcqmtx are the interaction indicators based on the propensity scores from the dichotomized strata, and anyhcq-anyivig are the indicators of treatment received.

reg totalVas anyivig34 anyhcq34 anymtx34 ivhc34 ivmtx34 hcmtx34 anyivig anyhcq anymtx

This is the implementation of the equation following table 2 (see table 10).

None of the coefficients is significantly different from 0, although the model shows a moderate relationship overall. We did not attempt to simplify this model, although one might omit the interactions and refit. If we omit the treatment indicators and perform the likelihood ratio test, we find a p-value of 0.046 suggesting a mild effect of treatment after adjustment for the strata. This could arise because we did not include some higher order interactions of the dichotomies and their interactions, or because the two highest strata were too broad a stratification.

Discussion

We have described two models for analysis of multiple therapies that are given in non-randomized fashion. This can be useful for research involving medical records or other observational studies. Two problems that arise here are missing data on the response variable and adjusting for the treatment given. For two treatment groups, drug vs. no drug, propensity score analysis has been widely used. However, when multiple drugs are given and possibly given in combination, the method of analysis is not clear. We have used two approaches. The first is to consider all one, two, and three drug combinations (one could also consider patients who received no drug, but that did not occur in this study), which leads to seven drug therapies. In our case, one combination had only one patient and was deleted from the analysis. After computing the multinomial logistic regression, we computed the predicted probability for each drug combination and used these as our propensity scores. We could stratify on these, but the analysis becomes cumbersome and difficult to interpret. Using these predicted probabilities as predictors for the severity of damage score is straightforward and has been used in the two group case.

The second method was to use the univariate propensity strata from the analyses comparing combinations of using each drug versus all other combinations that did not include the drug. We use the strata based on the upper 40% of the predicted probabilities of receiving the drug.

This gave us a set of dichotomous variables. We then used these and their two-way interactions to predict damage.

We found that the ability to predict damage was better using the multinomial logistic regression approach than with the combination of univariate strata approach. In both cases, the R^2 was modest ($R^2=0.20$ for the multiple logistic method and $R^2=0.14$ for the dichotomization method), as none of the variables predicted the damage well. Some candidate predictors had to be deleted as they were rare and did not contribute to prediction, or led to collinearity.

We did not use propensity score matching in these analyses, but are doing so in additional work on two-category models (e.g., anymtx vs. other). Using the multiple logistic predicted probabilities for treatment group leads to several issues. Should one match? And if so, how? If one selects the group with the largest score, a second group with almost the same score will be ignored. The problem is one of ordering in several dimensions. In such cases, we often map the scores to a single dimension. Thus, we decided to use the predicted risk for each category as a covariable.

Another alternative that we did not try is to use classification and regression trees (CART). At least one article (Murtagh, 1997) suggests it is not a useful approach in this case. All computations were done using Stata 10 (StataCorp). All presentations of data in this manuscript are preliminary.

Acknowledgments

We thank the International Myositis Assessment and Clinical Studies (IMACS) Group Members for contributing data to this study. We thank Drs. Michael Ward and Bo Zhen for critical reading of the manuscript. We thank the referees for their suggestions and careful reading of the manuscript.

This work was supported in part by the intramural research program of NIEHS, NIH, and by a research grant to Peter Lachenbruch from the Cure JM Foundation.

References

- Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer Til. Variable Selection for Propensity Score Models. *Am Jour Epid* 2006;163(12):1149–1156.
- D’Agostino, Ralph B, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 1998;17:2265–2281. [PubMed: 9802183]
- Isenberg DA, Allen E, Farewell V, Ehrenstein MR, Hanna MG, Lundberg IE, Oddis C, Pilkington C, Plotz P, Scott D, Vencovsky J, Cooper R, Rider L, Miller F. for the International Myositis and Clinical Studies Group (IMACS). International Consensus Measures for Patients with Idiopathic Inflammatory Myopathies. Development and Initial Validation of Myositis and Damage Indices in Patients with Adult Onset Disease. *Rheumatology* 42:1–7.
- Mendez EP, Lipton R, Ramsey-Goldman R, Roettcher P, Bowyer S, Dyer A, Pachman LM. NIAMS Juvenile DM Registry Physician Referral Group. US incidence of juvenile dermatomyositis, 1995–1998: results from the National Institute of Arthritis and Musculoskeletal and Skin Diseases Registry. *Arthritis Rheum* 2003;49(3):300–5. [PubMed: 12794783]
- Murtagh PW. Methods of Variable Selection in Regression Modelling *Commun. Statist –Simula* 1998;27(3):711–734.
- Rider LG, Miller FW. Idiopathic Inflammatory Muscle Disease: Clinical Aspects. *Balliere’s Clinical Rheumatology* 2000;14(1):37–54.
- Rider, LG.; Pachman, LM.; Miller, FW.; Bollar, H., editors. *Myositis and You*. Washington, DC: The Myositis Association; 2007.
- Rubin DB. Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Annals of Internal Medicine* 1997;127(8):757–763. [PubMed: 9382394]
- StataCorp. *Stata Statistical Software: Release 10*. College Station, TX: StataCorp LP; 2007.

Appendix 2: Stata commands

```
xi: logistic anyhcq age cardpul caucasian gender i. onsetseverity othrheum anyulc delaytodx
diag dysphagia if kidinclude==1, nolog
```

The “xi” instructs Stata to create indicator variables for any variable that is preceded by “i.” The logistic command asks for a logistic regression with dependent variable anyhcq. The “if” condition requires the variable kidinclude to be 1. This variable indicates that the patient is to be included in the analysis. The “nolog” option (options are always after a comma) tells Stata not to print the log of iterations. The variable cardpul indicates if cardiopulmonary involvement is present. The variable onsetseverity is coded in 4 levels ranging from mild to severe. The variable othrheum indicates if other rheumatic diseases are present. Anyulc indicates if there is a skin or gastric ulcer present. Delaytodx is the time in days from the current date to the date of diagnosis. Diag is 1 if dermatomyositis and 2 if polymyositis. We later considered models restricted to dermatomyositis. Dysphaagia indicates if the patient had difficulty swallowing.

The output showed that the overall model is highly significant ($p=0.0024$) and that some variables are of marginal importance. For example cardpul has a p-value of 0.780 and is not useful. Similarly, anyulcer was useful for prediction. We then issued the command “predict hat if kidinclude==1” to get predicted probabilities from this analysis.

We used the xtile command to create the strata which gives us the variable we need.

We find the effect of treatment after adjusting for the risk of getting hcq by an analysis of covariance. We note that the model is not a good predictor (the model has a low R^2), so even with a fairly strong ability to predict treatment choice, the strata and the treatment together do not provide much information on damage outcome.

The ANOVA command in Stata will perform an analysis of covariance by indicating if a covariate is continuous.

```
. anova totalVas strat anyhcq if kidinclude==1, cont(anyhcq) reg
```

The first argument of the ANOVA command is the response, the second is the factor variable and the third is a covariate. The order of the factor and covariates is unimportant. We treated anyhcq as a continuous variable, although since it is dichotomous, it could also be a factor. The largest category of strat is the comparison group. The overall ANOVA F was not significantly different from a zero effect. The individual comparisons were not significantly different from 0. The test of HCQ is a Wald test and is not significantly different from 0.

```
. estimates store a
```

The estimates store command saves data needed to compute a likelihood ratio test of a nested model. We now look at a simplified model that omits the strata. This gives a test of the usefulness of the strata.

```
. anova totalVas anyhcq if kidinclude==1 & e(sample)==1, cont(anyhcq) reg
```

Requiring $e(sample)=1$ ensures that the same observations are used in both analyses. The variable $e(sample)$ is set by Stata commands to indicate the observation was used in an analysis. When there are missing values on a variable, this allows us to conduct likelihood-ratio tests on the same set of data.

. lrtest a

This conducts the likelihood-ratio test of the usefulness of the strata. LR χ^2 with 4 d.f. = 4.85 with $p = 0.3029$. This assumes the models are nested.

Results for anymtx and anyvig are similar to the above and are not shown..

The command to compute the multivariate risk score uses multiple logistic regression. The command is

```
. xi:mlogit treat age caucasian gender i.onsev anyulc delaytodx dysphagia if kidinclude==1 &
treat~=5, nolog
```

The restriction statement (“if”) includes eligible children and omits patients with treatment =5.

Treatment 5 was the combination of hcq and ivig which had only 1 patient.

We obtain predicted scores as

```
. predict treat* if e(sample)
```

(option pr assumed; predicted probabilities)

(326 missing values generated)

The restriction if e(sample) constrains the predicted values to the estimation sample. This creates new variables treat1, treat2, ..., treat7 omitting an indicator for treatment 5.

The command to do the regression on the strata variables from the logit models is below.

```
. reg totalVas anyhc34 anymt34 anyiv34 hcmt hciv mtiv anyhcq anymtx anyvig if
kidinclude==1 & diag==1,
```

Table 1

Drugs used to treat Juvenile Dermatomyositis

Drug	Use	Side effects
Daily oral corticosteroids (Prednisone, methylprednisolone)	Always used to treat JM– improves weakness, rashes, and other manifestations	Weight gain, decreased growth, mood swings, weakness, high blood sugar, cataract, osteoporosis
IV pulse steroids (methylprednisolone)	More severe illness, may be used once or repeatedly	High blood sugar, blood pressure changes, facial flushing
Hydroxychloroquine (Plaquenil) [hcq]	For JDM skin rashes, may help in reducing steroid requirements	Digestive reactions, headache, retinal damage at high doses, increased skin rashes
Methotrexate (mtx)	Helps reduce steroid dose and duration, results in fewer steroid side effects (Ramanan et al 2005), affects most JM manifestations	Digestive reactions, mouth ulcers, liver toxicity, low blood counts, infection
Intravenous Immune Globulin (ivig)	Second-line agent, rapid control of JM manifestations, does not suppress immune system	Aseptic meningitis, headache
Cyclosporine	Second-line agent, improves muscle weakness, skin rashes, lung disease, fevers	High blood pressure, gingival hyperplasia, liver toxicity, tremors, kidney function
Azathioprine (Imuran) [Aza]	Second-line agent, improves muscle weakness and skin rashes, slower to see maximal effect, adult polymyositis use	Digestive reactions, liver toxicity, low blood counts, infections
Cyclophosphamide (Cytoxan)	Severe refractory disease, ulcerations, lung disease	Hemorrhagic cystitis, low white cell count, infections; infertility and malignancy with long-term use

Table 2

Abbreviations for Drug Propensity Score Strata

Drug	Indicator
Hydroxychloroquine (HCQ)	H60
Methotrexate (MTX)	M60
Intravenous immune globulin (IVIG)	I60

Table 3

Variable names

Variable (n)	Mean	Standard Deviation	Minimum	Maximum
Age (156)	9.16	4.05	1.65	17.75
Cardio pulmonary involvement (no=1, yes=2) (154)	1.01	0.11	1	2
Race- Caucasian (0=no/1=y) (156)	0.81	0.39	0	1
Gender (0=male/1=female) (154)	0.67	0.47	0	1
Cutaneous or gastrointestinal ulceration(0=no/1=yes) (154)	0.24	0.43	0	1
Other rheumatic disease(0=no/1=yes) (154)	0.032	0.18	0	1
Delay to diagnosis (days) (154)	204.43	450.62	0	3865
Dysphagia (0=no/1=yes) (150)	0.20	0.40	0	1
Diagnosis (1=dermatomyositis, 2=polymyositis) (156)	1.058	0.23	1	2
Onset severity (149) 4 Category	1 29.5% 2 42.3% 3 24.2% 4 4.0%		1	4
Onset severity (149) 3 category	1 29.5% 2 42.3% 3 28.2%		1	3

Table 4

Results of logistic regressions

Variable	OR HCQ	OR MTX	OR IVIG
Age	1.08	1.04	0.97
Cardio-pulmonary	1.50	0.38	1.36
Caucasian	0.52	1.74	1.26
Gender	1.40	1.67	1.79
Onset severity 2	0.30**	1.17	3.12
Onset severity 3	0.17***	2.09	6.96*
Onset severity 4	0.06*	4.22	4.94
Other rheumatic disease	0.06	0.78	not used – perfect prediction
Any ulceration	1.08	1.08	2.47
Delay to diagnosis	1.0007	1.005	0.9999
Diagnosis	0.14	0.81	0.51
Dysphagia	2.08	0.17	1.23

Notes: OR – Odds Ratio, HCQ – hydroxychloroquine, MTX – methotrexate, IVIG – intravenous immune globulin

Table 5

Analysis of Variance for totalVas (Severity of Damage) using strata of risk of receiving HCQ

N=131	F (5, 125) = 1.38		p=0.24	RMSE=5.54
Source	Coefficient	S.E.	95% confidence interval	
_cons stratum	2.74	1.38	0.0017	5.47
1	1.18	1.67	-2.13	4.50
2	0.54	1.63	-2.69	3.77
3	2.29	1.54	-.76	5.34
4	-.83	1.54	-3.88	2.23
5	(dropped)			
anyhcq	1.76	1.08	-.37	3.90

Table 6

Number of drugs excluding Prednisone

Number	0	1	2	3	4	5+
All drugs	14	31	33	31	24	20
HCQ, MTX, IVIG	31	51	49	22	-	-

Table 7

Propensity Score Variable Names

Treatment	Propensity Score
anyhcq	Treathat1
Anita	Treathat2
anyivig	Treathat3
anyhcq and anymtx	Treathat4
anyhcq and anyivig	Omitted – only 1 patient
anymtx and anyivig	Treathat6
anyhcq, anymtx and anyivig	Treathat7

Table 8

Multiple linear regression of totalVas on propensity scores and treatment indicators

Variable	Coef.	Std. Err.	t	P
Treathat1	3.66	5.27	0.69	0.49
Treathat2	−12.61	6.21	−2.03	0.045
Treathat3	13.95	7.98	1.75	0.084
Treathat4	−0.15	3.42	−0.04	0.97
Treathat6	4.85	4.42	1.10	0.28
Treathat7	16.31	5.97	2.73	0.008
Anyhcq	1.07	1.29	0.83	0.41
Anymtx	2.24	1.40	1.60	0.11
anyivig	−0.67	1.34	−0.50	0.62

Table 9

Results of logistic regression of anymtx on anyhcq, anyivig and their interaction

Variable	Coef.	Std. Err.	Z	P
Anyhcq	0.83	0.40	2.07	0.038
Anyivig	0.83	0.50	1.67	0.095
Anyhcq*Anyivig	1.68	1.17	1.43	0.15

Table 10

Regression of totalVas on Stratum Predictors and interactions

N = 129	F(9, 119) = 2.10	p-value 0.0349	R² = 0.1369	Root MSE = 5.6403
Variable	Coef.	Std. Err.	T	P
anyhc34	-1.66	1.40	-1.19	0.24
anymt34	-0.64	2.33	-0.28	0.78
anyiv34	-1.04	1.89	-0.55	0.58
hcmt	2.06	2.74	0.75	0.45
hciv	1.88	3.09	0.61	0.54
mtiv	3.51	2.74	1.28	0.20
anyhcq	1.21	1.12	1.08	0.28
anymtx	1.44	1.15	1.25	0.21
anyivig	1.67	1.21	1.38	0.17