

Supplementary file

Table S1. The PICO database search strategy

| PICO component | Search term |
|----------------|--|
| Patient | 1. (\$arthrit\$ or rheum\$).mp. |
| Intervention | 2. (\$dmard or methotrexate or disease modifying or antirheum\$ or MTX or disease-modifying).mp. |
| Comparator | Not/Applicable |
| Outcome | 3. medic\$ possession ratio.mp. 4. ((claim\$ adj3 medic\$) or (claim\$ adj3 treatment) or (insurance adj3 medic\$) or (insurance adj3 health) or Medicaid).mp. 5. (\$adher\$ or complian\$ or (refusal adj3 medic\$) or (refusal adj3 treatment) or (refusal adj3 patient)).mp. 6. ((\$concord\$ adj3 medic\$) or (\$concord\$ adj3 treatment) or (\$concord\$ adj3 patient)).mp. 7. (missed medi\$ or pharmacy refill or (resist\$ adj3 medic\$) or (resist\$ adj3 treatment) or (resist\$ adj3 patient)).mp. 8. ((discontin\$ adj3 medic\$) or (discontin\$ adj3 treatment) or (discontin\$ adj3 patient) or (persist\$ adj3 medic\$) or (persist\$ adj3 treatment) or (persist\$ adj3 patient)).mp. 9. 3 or 4 or 5 or 6 or 7 or 8 10. 1 and 2 and 9 11. remove duplicates from 10 |

mp = multiple purpose (Title, Original Title, Abstract, Subject Heading, Name of Substance, and Registry Word fields)

\$=wildcard adj3= Search term occurs within three or fewer words of specific term

Appendix B

Table S2. The Quality Assessment (QA) Tool used to assess the quality of included articles.

| Quality assessed | Coding framework | | |
|---|------------------|-------------------------|---------------------------|
| Appropriate methods to select participants | | | |
| 1 Sampling frame, age and sex of sample described. | yes | no | don't know |
| 2 >80% participation or comparison of consents and refusals | yes | no | don't know |
| Appropriate methods to measure adherence | | | |
| 3 Measure of adherence reproducible | yes | no | don't know |
| 4 Adherence Measure | Objective | Validated subjective | Unvalidated subjective |
| Appropriate Measure of outcome | | | |
| 5* Measure of outcome/predictor reproducible | yes | no | don't know |
| 6* Objective measure of outcome/predictor | yes | no | don't know |
| Appropriate methods to reduce bias | | | |
| 7 Recall period less than 2 weeks | yes | no | don't know |
| 8 Consecutive, stratified, or randomised sample | yes | no | don't know |
| Appropriate methods to reduce confounding | | | |
| 9* Confounding controlled for in analysis | yes | no | don't know |
| 10* Confounding quantified in analysis | yes | no | don't know |
| Appropriate methods to report adherence statistics | | | |
| 11 Adherence reported with a measure of dist. | yes | no | don't know |
| Appropriate methods to analyse strength of association with outcome | | | |
| 12* Association reported with 95% CI | yes | no | don't know |
| 13* Association accounts for skewedness | yes | no | don't know |
| 14 N>30 = 1 | yes | no | don't know |
| 15* 10 cases per predictor | yes | no | don't know |
| Conflict of interest | | | |
| 16 Conflict of Interest declaration | yes | no | don't know |

Table S3. QA scoring for each article included in systematic review

| Author | QA tool item number | | | | | | | | | | | | | | | | Score |
|--|---------------------|---|---|---|-----|-----|---|---|-----|-----|----|-----|-----|----|-----|----|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | |
| Cannon et al. (2011) ²⁸ | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 15/17 |
| Grijalva et al. (2007) ²³ | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 15/17 |
| Waimann et al. (2013) ³⁰ | 0 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 15/17 |
| de Thurah et al. (2010) ²⁶ | 0 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 14/17 |
| de Thurah et al. (2010) ²⁷ | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 14/17 |
| de Klerk et al. (2003) ²² | 0 | 0 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 12/17 |
| Contreras-Yanez al. (2010) ²⁴ | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 11/17 |
| Salt & Frazier (2011) ²⁹ | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 9/17 |
| Grijalva et al. (2010) ^{25*} | 1 | 1 | 1 | 2 | n/a | n/a | 1 | 1 | n/a | n/a | 1 | n/a | n/a | 1 | n/a | 0 | 9/10 |
| Harley, Frytak & Tandon(2003) ^{21*} | 1 | 1 | 1 | 2 | n/a | n/a | 1 | 1 | n/a | n/a | 0 | n/a | n/a | 1 | n/a | 0 | 8/10 |

(*These studies did not investigate an association with adherence to MTX and were evaluated using the ten point scale)

Table S4. Summary of evidence for the association of adherence to MTX with clinical and patient reported outcomes

| Study | Predictor | Outcome | Unadjusted Effect | <i>P</i> | Adjusted Effect Size | <i>P</i> |
|-----------------|------------------------|-----------------------------------|--------------------------------|----------|-----------------------------|----------|
| Contreras- | Nonadherent CQ ≤8 vs | Incidence of disease flare (n=14) | 48.41 per 100 vs 13.31 per 100 | 0.0 | NP RR 4.8 (0.8, 27.6) | NP 0.08 |
| Cannon et al. | | | | | | |
| Full cohort | MPR ≥ 80% | Mean Difference in | β -0.34 (-0.68, -0.06) < | < | β -0.37 (-0.67,-0.07) | < 0.05 |
| First-time user | MPR ≥ 80% | Mean Difference in | β -0.54 (-1.18, 0.11) NS | NS | β -0.40 (-1.11, 0.30) | NS |
| | MPR <80% (17) vs MPR ≥ | Erythrocyte | 35.2±26 vs 25±17 | 0.0 | NP | NP |
| | | C-Reactive Protein | 1.8±1.6 vs 1.4±1.9 | 0.5 | NP | NP |
| | | Tender joint count | 4.9±4.7 vs 3.6±4.4 | 0.2 | NP | NP |
| | | Swollen joint count | 4.6±4.6 vs 3.2±3.5 | 0.1 | NP | NP |
| | | Health Assessment | 1.0±0.45 vs 0.9±0.54 | 0.6 | NP | NP |
| | | Visual Analogue Scale | 37±21 vs 32±19 | 0.2 | NP | NP |
| | | Patient global health | 41±23 vs 41±21 | 0.9 | NP | NP |
| | | Pain | 3.8±2.9 vs 4.4±2.1 | 0.3 | NP | NP |
| Established | MPR ≥ 80% | Mean Difference in | β -0.38 (-0.67, -0.05) < | < | β -0.37 (-0.72,-0.02) | < 0.05 |
| | MPR<80% (n=54) vs MPR | Erythrocyte | 27.4±23.0 vs | 0.2 | NP | NP |
| | | C-Reactive Protein | 1.6±1.5 vs 1.2±1.2 | 0.0 | NP | NP |
| | | Tender joint count | 5.6±6.7 vs 3.7±4.6 | 0.0 | NP | NP |
| | | Swollen joint count | 4.3±4.9 vs 3.2±3.2 | 0.0 | NP | NP |
| | | Health Assessment | 0.9±0.6 vs 0.9±0.9 | 0.7 | NP | NP |
| | | Visual Analogue Scale | 35±28 vs 39±20 | 0.2 | NP | NP |
| | | Patient global health | 34±24 vs 33±17 | 0.4 | NP | NP |
| | | Pain | 4.5±2.6 vs 4.2±2.2 | 0.3 | NP | NP |
| Waimann et al. | % Adherence (MEMS) | DAS28 at 2 years | β -0.02 (-0.03, - | NP | β -0.2 | 0.03 |
| | <80% (n=61) vs ≥ | Sharp score at baseline | 58 vs 80 | 0.0 | NP | NP |
| | <80% (n=45) vs ≥ | Sharp score at 12 | 61 vs 86 | 0.0 | NP | NP |
| | <80% (n=41) vs ≥ | Sharp score at 24 months | 69 vs 87 | 0.1 | NP | NP |
| | <80% (n=84) vs ≥ | M-HAQ at baseline | 0.92 vs 0.66 | 0.0 | NP | NP |
| | <80% (n=78) vs ≥ | M-HAQ at 12 months | 0.74 vs 0.69 | NS | NP | NP |
| | <80% (n=78) vs ≥ | M-HAQ at 24 months | 0.82 vs 0.74 | NS | NP | NP |

* Studies judged low quality; ^a MTX only; CQ = Compliance Questionnaire; MPR = Medication Possession Ratio; MEMS = Medicine Event Monitoring System; M-HAQ= Modified Health Activity Questionnaire; RR = Relative risk; MD= mean difference; β = regression coefficient NS = non-significant; NP = not presented.