

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

REAL WORLD EFFECTIVENESS OF BARICITINIB IN THE SWISS RHEUMATOID ARTHRITIS REGISTER (SCQM-RA)

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2023-072300
Article Type:	Original research
Date Submitted by the Author:	03-Feb-2023
Complete List of Authors:	Gilbert, Benoit Thomas P.; Geneva University Hospitals, Division of Rheumatology, Department of Internal Medicine Specialties Mongin, Denis; Hopitaux Universitaires de Geneve, Rheumatology, Department of Internal Medicine Specialties Aymon, Romain; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties Lauper, Kim; Hopitaux Universitaires de Geneve, Rheumatology, Department of Internal Medicine Specialties LAEDERMANN, Cédric; Eli Lilly PERRIER, Clémentine; Eli Lilly Mueller, Ruediger; Rheumazentrum Ostschweiz, private practice Courvoisier, Delphine; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties Finckh, Axel; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties
Keywords:	RHEUMATOLOGY, EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

REAL WORLD EFFECTIVENESS OF BARICITINIB IN THE SWISS RHEUMATOID ARTHRITIS REGISTER (SCQM-RA)

Benoit GILBERT^{1*}, Denis MONGIN¹, Romain AYMONT¹, Kim LAUPER^{1,2}, Cédric LAEDERMANN³,
Clémentine PERRIER³, Ruediger MUELLER⁴, Delphine S. COURVOISIER¹, Axel FINCKH¹

¹ Division of Rheumatology, Geneva University Hospitals, Geneva.

² Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research,
University of Manchester.

³ Eli Lilly (Suisse) SA, 16 Ch Des Coquelicots, CH-1214 Vernier.

⁴ Private practice, St Gallen, Switzerland.

* Correspondence:

Benoît Thomas P. GILBERT

26 Avenue de Beau-Séjour

Division of Rheumatology, Department of Medicine, Geneva University Hospital,

1206 Geneva,

Switzerland

benoit.gilbert@etu.unige.ch

ABSTRACT

Objectives: This observational study compares the effectiveness of baricitinib (BARI), a targeted synthetic DMARD (tsDMARD) with alternative biological DMARDs (bDMARDs) in rheumatoid arthritis (RA) patients, from a prospective, longitudinal cohort.

Methods: We compared patients initiating a treatment course of BARI, tumor necrosis factor inhibitors (TNFi) or other mode of action bDMARDs (OMA), during a period when all these DMARDs were available in Switzerland. The primary outcome was drug-maintenance; secondary outcomes included discontinuation rates related specifically to ineffectiveness and to adverse events. We further analyzed rates of low disease activity (LDA) and remission (REM) at 12 months, and drug maintenance in b- and tsDMARD-naïve population.

Results: A total of 1053 treatment courses (TC) were included: 273 on BARI, 473 on TNFi and 307 on OMA. BARI was prescribed to older patients with longer disease duration and more previous treatment failures than TNFi. Compared to BARI, the adjusted drug maintenance was significantly shorter for TNFi (hazard ratio (HR) for discontinuation: 1.76; 95% CI [1.32-2.35]), but not compared to OMA (HR 1.27; 95% CI [0.93-1.72]). These results were similar in the b/tsDMARD-naïve population. The higher discontinuation of TNFi was mostly due to an increased discontinuation for ineffectiveness (HR = 1.49; 95% CI [1.03 – 2.15]), with no significant differences in drug discontinuation for adverse events (HR = 1.46; 95% CI [0.83 - 2.57]). The LDA and REM rates at 12 months did not differ significantly between the 3 groups.

Conclusions: BARI demonstrated a significantly higher drug maintenance compared to TNFi, mainly due to lower drug discontinuations for ineffectiveness, but similar maintenance to OMA. Clinical outcomes did not differ between the three groups. Our results suggest that BARI is an appropriated therapeutic alternative to bDMARDs in the management of RA.

Strengths and limitations of this study

Strengths:

- Use data derived from office-based rheumatologists
- Study period where all alternative medications were available on the market
- Several sensitivity analyses, congruent with main results

Limitations:

- Not a randomized setting
- Sub-analysis in b/tsDMARD-naïve population has limited sample size

INTRODUCTION

Rheumatoid arthritis (RA) is an auto-immune disease leading to widespread inflammation and irreversible joint damage, if insufficiently treated. New treatment paradigms have emerged in the last decades, such as “early aggressive therapy” in the so called “window of opportunity”, during which patients are more likely to reach long term remission.[1] A wide panel of biological disease modifying antirheumatic drugs (bDMARDs) and targeted synthetic DMARDs (tsDMARDs) have been approved in the management of RA, after failure of methotrexate. In clinical-trial settings, b- and tsDMARDs have demonstrated significant reduction of joint inflammation and prevention of joint damage.[2–8]

Efficacy estimates from placebo-controlled randomized trials often differ from real-world effectiveness estimates, because of patient selection, adherence to therapy and other reasons.[9–12] Indeed, drug maintenance of many bDMARDs remains modest in observational analyses, while long term remissions are rare and secondary loss of efficacy frequent.[13] Furthermore, understanding the clinical effectiveness of bDMARDs or tsDMARDs in specific conditions, such as elderly or multi-morbid patients, may become important as we move towards personalized care. Finally, trials provide only limited data on long term effectiveness and safety because clinical-trial follow-up is typically less than 12 months.

Baricitinib (BARI) has been approved in Switzerland for the treatment of RA in 2017 as well as all around the world. Clinical trials with BARI have established efficacy and demonstrated acceptable adverse events profile, both in combination with methotrexate or in monotherapy.[14–20] However, evidence about effectiveness of BARI compared to TNFi in

1
2
3 real-world settings are scarce. A recently published analysis of registry data from Sweden
4
5 showed that baricitinib had higher maintenance as compared to most other bDMARD.[21]
6
7
8 Pappas et al., in the United States, also demonstrated that TNFi and non-TNFi drugs had
9
10 similar outcomes when prescribed in b/tsDMARD-naïve population, an observation replicated
11
12
13 in the RA-BE-REAL study.[22,23]
14

15 The aim of our analysis was to compare real-world drug maintenance between BARI and other
16
17 approved b/tsDMARDs, using data from a European registry.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODS

Study population

This is a nested cohort study from a prospective, longitudinal, cohort of Swiss RA patients in a real-life setting, the Swiss Clinical Quality Management registry (SCQM). The SCQM registry was founded in 1997 with the financial support of Swiss regulatory authorities, who recommended a continuous monitoring of all patients receiving new DMARDs. Unlike many other European registries, most patients are enrolled by private office-based rheumatologists (60%), providing a population-based sample of RA patients in Switzerland. All approved RA treatments are represented in the registry. The data for this analysis was extracted from the SCQM registry on 2020-06-01.

We used “treatment courses” (TCs) as our denominator of interest, with each new treatment initiation considered as a separate “treatment course” (TC). We included all TCs with the medications of interest initiated between 2017-09-01 and 2020-06-01, with at least one follow-up visit, in adult patients with a diagnosis of RA confirmed by a rheumatologist. Thus, a given patient could potentially contribute to several TCs during the study period. To minimize the risk of confounding bias, the time window was selected to include only the period when all the therapies examined were available for prescription and reimbursed (BARI was first reimbursed on the Swiss market in September 2017). We excluded TCs with no follow-up visit at the time of data extraction.

Exposure of interest

The exposure of interest was the type of treatment used, namely BARI, TNFi, and other mode of action bDMARDs (OMA), excluding other tsDMARDs and rituximab. We decided to exclude rituximab a priori, because its long-term action impairs precise estimation of treatment discontinuation. Tofacitinib was excluded because we had insufficient TC to perform meaningful comparative effectiveness analyses against a single other specific tsDMARD agent. Included TNFi treatments were: adalimumab, etanercept, golimumab, certolizumab, infliximab. Included OMA treatments were: tocilizumab, abatacept, sarilumab, and anakinra.

Outcomes

The primary outcome of this analysis was the time to all-cause-discontinuation. This outcome, also referred to as “drug maintenance”, captures both the drug’s effectiveness and its tolerance.[24] The time to all-cause-discontinuation was defined as the number of days between treatment initiation and the reported date of discontinuation, or the date of initiation of a new b/tsDMARD, whatever came first. In survival analyses, death or lost-to-follow up are censored. We also report discontinuation rates at 12 months. Temporary discontinuations of less than 6 months (for instance, because of an elective surgery or a pregnancy) were not considered a permanent drug discontinuation. Discontinuation reasons are recorded by the clinician when stopping a DMARD treatment, who chooses between four options (“Adverse event”, “Ineffectiveness”, “Remission”, or “Other”).

Pre-planned secondary outcomes, were time to discontinuation due to ineffectiveness and time to discontinuation due to adverse events. Other secondary outcomes included response rates, namely the rates of low disease activity (LDA) and remission (REM), at 12 months, defined respectively as attaining a CDAI score ≤ 10 and CDAI score ≤ 2.8 (not mutually

1
2
3 exclusive).[25] Finally, we performed an exploratory subgroup analysis, restricting the
4
5 population to b/tsDMARD-naive patients only, and re-assessing the main outcome in this
6
7 setting.
8
9

10 11 12 13 **Statistical analysis** 14

15 Analyses were conducted and reported in accordance to EULAR recommendations for
16
17 comparative effectiveness research.[9] Baseline characteristics were compared using ANOVA
18
19 or χ^2 tests as appropriate. For the primary outcomes, Kaplan-Meier survival analyses were
20
21 used to assess crude drug maintenance, and groups were compared using Log-Rank tests.
22
23 Subsequently, missing covariates were imputed using chained equations (see below for
24
25 details). We then implemented Cox proportional hazard ratio models, to obtain adjusted
26
27 estimates. Based on prior subject matter knowledge,[26] we adjusted our models for the
28
29 following potential confounders: age, gender, BMI, concomitant csDMARDs use (yes/no),
30
31 concomitant prednisone usage (yes/no), CDAI score at baseline, disease duration, smoking
32
33 status (current-, former-, never-smoker) and line of therapy (1st, 2nd, 3rd, 4th or more), and
34
35 seropositivity (yes/no). Detailed definitions for each variable are available in the supplement.
36
37 A cluster term accounted for patients with multiple TCs. All conditions of application of the
38
39 Cox model were verified. One additional sensitivity analysis was conducted for the primary
40
41 outcome, using augmented inverse probability of treatment weighting (AIPW).
42
43
44
45
46
47
48
49
50
51

52 In secondary analyses, we used the Fine-Gray approach to assess specific reasons for drug
53
54 discontinuation (i.e. ineffectiveness, or adverse event) in a competing-risk setting. Other
55
56 secondary outcomes included response rates (LDA and REM) at 12 months. To avoid
57
58 overestimations, we computed the response rates using the 'confounder-adjusted response
59
60

1
2
3 rate with attrition correction' (CARRAC) method.[27] The latter estimates the response rates
4
5 using multiple imputations, with a model including both confounders and treatment stop
6
7 reason. CARRAC thus provides reliable estimates when reasons for treatment discontinuation
8
9 differ between compared groups.
10
11

12
13 For all adjusted analyses, missing baseline covariates were imputed using closest value in a
14
15 window of -90 to +30 days. However, this window was reduced to -30 to +7 days when
16
17 imputing baseline CDAI. If still missing after this first step, baseline CDAI values were imputed
18
19 using linear mixed effect regression model with quadratic time. We imputed other baseline
20
21 covariates with chained equations technique, which provides unbiased estimates if the
22
23 variables are missing at random.[27] Such imputations were performed using 50 datasets with
24
25 25 iterations. Imputation was done using the whole data set, before adequately subsetting
26
27 the data for each group comparison.
28
29
30

31
32 We also imputed data required for secondary outcomes, including disease activity. If the CDAI
33
34 score at 12-month was not available, the closest value in a window of +/- 45 days was used (a
35
36 3-month-wide window). If still missing, the 12-month CDAI values were imputed using nearest
37
38 neighboring value, as previously described.[28]
39
40

41
42 All analyses were conducted using R (version 4.0.3), in particular with packages "*tableone*",
43
44 "*survival*", "*mice*".[29] Two tailed p-values < 0.05 were considered significant. We did not
45
46 adjust p value for multiple comparisons, as outcomes were pre-specified. Final analysis code
47
48 is shown in the supplement (SUPP_R_CODE).
49
50

51 **Patient and Public Involvement**

52
53 Patient involvement is central to the SCQM cohort. Several patients are part of the executive
54
55 board and involved in the approval of research projects.
56
57
58
59
60

RESULTS

Population description

During the study period, 1053 TC were initiated in 834 different patients, including 273 TC with BARI, 473 with TNFi and 307 with OMA (Figure 1). TNFi were more often given as a second line therapy after methotrexate failure. Inversely, BARI was prescribed to significantly older patients, with longer disease durations and more previous treatment failures (Table 1).

For peer review only

Table 1: Baseline characteristics of study population, SCQM-RA registry, 2017-2020.

Variable	BARI (TC = 273; 273 pts)		TNFi (TC = 473; 408 pts)		OMA (TC = 307; 289 pts)		p values
	n % of total in group Otherwise: Mean (SD)						
Patients		Miss.		Miss.		Miss.	
Female	78 %	0	74 %	1	73 %	1	0.28
Age (years)	59 (14)	0	52 (15)	1	59 (13)	1	<0.01
Disease duration (years)	13 (10)	4	8 (9)	19	11 (9)	5	<0.01
CDAI baseline (raw data)	19 (10)	175	18 (10)	301	20 (13)	204	0.25
CDAI baseline (imputed)	15 (9)	0	14 (9)	0	16 (11)	0	0.05
Obesity (BMI > 30)	16 %	104	14 %	134	13 %	115	0.59
Smoking	17 %		18 %		21 %		
Current	28 %	32	26 %	69	28 %	26	0.90
Former	43 %		41 %		48 %		
Never							
Seropositive (ACPA or RF)	75 %	1	70 %	7	77 %	5	0.07
TC		Miss.		Miss.		Miss.	
Concomitant csDMARD	40 %	0	61 %	0	46 %	0	<0.01
Line of Therapy	17 %		48 %		22 %		
1 st (= bio-naïve)	20 %	0	23 %	0	24 %	0	<0.01
2 nd	19 %		11 %		24 %		
3 rd	44 %		18 %		31 %		
4 th or later							
Previous tsDMARD use (non-BARI)	33 %	0	1 %	0	5 %	0	<0.01
Concomitant glucocorticoid (at any time)	22 %	0	20 %	0	24 %	0	0.28
Mean dose of concomitant glucocorticoid (mg)	2.0 (4.6)	0	2.1 (5.4)	0	2.2 (5.1)	0	0.90
Dose of BARI (4mg)	86 %	0	-	-	-	-	-

Table 1 Legend: In Switzerland, BARI was prescribed to older patients, with longer disease duration and more previous treatment failures. Missing values for covariables are reported as absolute numbers
 BARI = baricitinib. TC = Treatment Courses. CDAI = Clinical Disease Activity Index. TNFi = Tumor Necrosis Factor Inhibitors. OMA = bDMARDs with Other Mode of Action. tsDMARD = targeted synthetic DMARDs. ACPA = Anti Citrullinated Peptide Antibodies. RF = Rheumatoid Factor. Miss. = number of missing values. P-values are

1
2
3 obtained by 3-way ANOVA or Chi2. In TFNi and OMA groups, some patients have contributed several TC, thus
4 total number of TCs exceeds total number of patients.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Time to all-cause-discontinuation

Table 2: Crude treatment discontinuation by group and by reason, SCQM, 2017-2020.

	BARI (TCs = 273)	TNFi (TCs = 473)	OMA (TCs = 307)
Treatment discontinuation (all causes)	30 %	43 %	35 %
For adverse events	8 %	10 %	10 %
For ineffectiveness	16 %	23 %	17 %
For remission	0 %	1 %	0 %
For other reason	5 %	8 %	7 %

Table 2 legend: % are computed on total number of TCs per group, for the whole study period.
BARI = baricitinib. TNFi = TNF inhibitors. OMA = Other Mode of Action drugs. TC = Treatment Courses. Due to rounding, the sum of the percentages of the causes of discontinuation may not correspond exactly to the total treatment discontinuation percentage.

Crude proportions of treatment discontinuation by reasons are reported in Table 2, and crude times of observation are represented on Figure S1.

At 12 months, based on the Kaplan-Meier curves (Figure 2), the estimated proportions of patients still on therapy were: 71% (95% CI [65% - 77%]) in the BARI group, 55% (95% CI [50% - 61%]) in the TNFi group, and 63% (95% CI [57% - 70%]) in the OMA group.

Overall, unadjusted time to all-cause-discontinuation was significantly longer in the BARI group compared to the TNFi group (estimated median prescription survival-time of 704 versus 448 days; Log-rank $p < 0.01$; Figure 2). These results persisted after adjustment for confounding factors using the multivariable Cox model (HR = 1.76; 95% CI [1.32-2.35]; $p < 0.001$; Table S1 and Figure S2; Figure S3).

BARI versus OMA time to all-cause-discontinuation was not significantly different, even after adjustment (HR 1.27; 95% CI [0.93-1.72]; $p = 0.13$; Table S1, Figure S2 and Figure S3).

1
2
3 Sensitivity analyses using AIPTW led to similar conclusions (Figure S4). Covariates significantly
4 associated with decreased drug maintenance were high baseline CDAI scores and concomitant
5 glucocorticoid usage (Table S1 and Figure S2).
6
7
8
9

10 **Time to all-cause-discontinuation in b/tsDMARD-naïve patients**

11
12
13 In this exploratory subgroup analysis, we restricted the population to patients without prior
14 experience of b/tsDMARDs (so-called 'bio-naïve' patients, i.e. first b/tsDMARD prescription
15 after methotrexate failure). In this subpopulation, patient characteristics were more
16 balanced than in the main analysis, except for age, which remained younger in TNFi
17 population, and concomitant csDMARDs usage (more frequent in TNFi) (Table S2). Of note,
18 the sample size was consequently reduced to 46 BARI, 225 TNFi and 66 OMA.
19
20
21
22
23
24
25
26

27 When analysing only these b/tsDMARD-naïve patients, both the non-adjusted (Figure 3) and
28 the adjusted differences between BARI and TNFi became larger (HR TNFi vs BARI = 2.5; 95% CI
29 [1.23 – 5.16]; p=0.01), but the differences between baricitinib and OMA group remained not
30 significantly different (HR OMA vs BARI = 1.90; 95% CI [0.71 – 5.1]; p=0.2).
31
32
33
34
35
36

37 **Time to discontinuation for adverse events or ineffectiveness**

38
39
40 A secondary outcome was the cumulative incidence of drug discontinuation by specific
41 reasons for discontinuation (ineffectiveness or adverse events, Figure 4). Using Fine-Gray
42 adjusted approach, we found no difference in the incidence of adverse event comparing BARI
43 to TNFi (HR = 1.46; 95% CI [0.83 - 2.57]; p=0.13), or BARI to OMA (HR = 1.34; 95% CI [0.74 -
44 2.42]; p=0.25). The incidence of drug discontinuation for ineffectiveness was more frequent
45 in TNFi compared to BARI (HR = 1.49; 95% CI [1.03 – 2.15]; p=0.01), but similar between OMA
46 and BARI (HR = 1.09; 95% CI [0.72 – 1.64]; p=0.69).
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Remission and low disease activity at 12 months

The estimated 12-month rates of REM and LDA, estimated using CARRAC did not differ significantly between the 3 groups (Figure 5). LDA ranged from 62% to 71% and REM ranged from 17% to 26%.

DISCUSSION

In this study, the overall drug maintenance of BARI was significantly longer compared to TNFi, despite the fact that it was prescribed to older patients, with longer disease duration, and more previous treatment failures similar to what was observed in RA-BE-REAL, another real-world study.[23] However, the adjusted 12-month response rates in terms of LDA and REM did not differ significantly between BARI, TNFi and OMA groups. The difference in drug discontinuation owes mainly to more treatment discontinuations for ineffectiveness in the TNFi group compared to the BARI group, while drug discontinuation due to adverse event did not differ significantly between the groups.

Our results are in line with previous findings comparing other JAK-inhibitors (JAKi) (i.e. tofacitinib as well as BARI) to TNFi and OMA medications,[22,30] which reported a longer drug maintenance of tsDMARD compared to TNFi, and similar maintenance to other bDMARDs. Of note, Lauper et al., using data from 19 national registers, found no difference in retention time between JAK-inhibitors and TNFi.[31] Still, Lauper et al. grouped all JAKi together in their study, thus it is not clear if these observations remain true for BARI alone, which might differ from other JAKi. For instance, Barbulescu et al. reported a higher drug maintenance for BARI as compared to tofacitinib.[21]

It was previously shown that BARI is more efficient in relieving pain as compared to adalimumab therapy [32] and some molecular mechanisms relevant to JAK-STAT signalling

1
2
3 have been hypothesized.[33] This observation has been hypothesised to result anti-
4
5 nociceptive effect independent from inflammation.[33] This faster pain relief could partially
6
7 explain why BARI has increased maintenance than other medication in our study, even though
8
9 having similar 12-months LDA and REM rates. An alternative hypothesis is that the more
10
11 convenient oral administration encourages patients to stay on medication longer. Yet, a third
12
13 possible interpretation is that patients who experienced numerous treatment failures tend to
14
15 stay on their latest therapy; however, our study accounts for this potential bias, by performing
16
17 a sensitivity analysis in a subgroup of b-tsDMARD naïve patients, which showed a similar
18
19 result. Finally, given recent discussion regarding tofacitinib safety,[34] future research needs
20
21 to clarify whether a class effect for JAKi related adverse events exist. In this analysis, we found
22
23 no indication for an increased incidence of adverse-related treatment discontinuation with
24
25 BARI compared to alternative bDMARDs. Randomized controlled trials are ongoing to further
26
27 compare safety profile of BARI versus TNFi (NCT04086745 and NCT03915964).
28
29
30
31
32
33

34 35 **Limitations and Strengths**

36
37 This work has several limitations, mostly inherent to the observational setting. First, as this is
38
39 a non-randomized study, we cannot formally exclude unmeasured confounding between the
40
41 groups. The available baseline variables were, in most cases, adequately balanced, except for
42
43 age. When we restricted the analysis to the subgroup of b/tsDMARD naïve patients, we found
44
45 largely similar results. Despite being limited by the small sample size, this exploratory
46
47 subgroup analysis suggests that confounding by line of treatment was adequately accounted
48
49 for in the adjusted analysis.
50
51
52

53
54 Secondly, the average length of follow-up was only approximatively 200 days per TC (Figure
55
56 S1). Indeed, our study covers about 2 and a half years, and we only included TC newly initiated
57
58 during this time-windows. Also, because of the study setting, as much as 65% of TC did not
59
60

1
2
3 have CDAI scores recorded at the date of initiation, and many were missing at the 12-month
4 exact timepoint (Figure S5). Hence, our analysis of response rates relied heavily on linear
5 interpolation techniques, using other available timepoints, which results in large confidence
6 intervals for estimated response rates.[28]
7
8
9
10
11

12 The main strength of the study is that it relies on real-world data, and includes a relatively
13 large number of patients. As these patients are mostly treated by office-based
14 rheumatologists, our study population is representative of routine clinical practice. Also, sub-
15 group analyses and sensitivity analyses were consistent with the main results.
16
17
18
19
20
21

22 **CONCLUSIONS**

23
24
25 In this non-randomized prospective cohort study, we demonstrate that treatment with BARI
26 has at least similar effectiveness outcomes as alternative bDMARDs. Based on available data,
27 the estimated 12-month response rates did not significantly differ between BARI, TNFi and
28 OMA groups. We found no difference in treatment discontinuation for adverse event between
29 the three groups. Overall, our results are in line with findings from randomized trials, confirm
30 the effectiveness of BARI in daily practice and validate this agent as an alternative to bDMARDs
31 in RA.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

OTHERS

Contributorship statement

BG: data-management, data-analysis, figures, manuscript drafting.

DM: data-management, data-analysis, figures, manuscript-revision.

RA: data-analysis (in particular, sensitivity analyses), manuscript revision.

KL: data-analysis, manuscript revision.

CL: study design, manuscript revision.

CP: study design, manuscript revision.

RM: study design, manuscript revision.

DC: study design, data-analysis and interpretation, manuscript revision.

AF: study design (principal investigator), data analysis and interpretation, manuscript revision.

Competing interests

Benoît GILBERT has been once a paid speaker (Eli Lilly) and participated in advisory board (Janssen). Clementine PERRIER is employed by Eli Lilly and holds stock options (Eli Lilly and Company). Cedric LAEDERMANN is employed by Eli Lilly and holds stock options (Eli Lilly and Novartis). Axel FINCKH has received grants or contracts (Eli-Lilly, Pfizer, Abbvie, Gilead, BMS), consulting fees (Astra-Zeneca, Abbvie, Pfizer, Gilead), honorary payments (BMIS, Abbvie, Eli Lilly, Pfizer, MSD), and participated in advisory boards (Astra-Zeneca, Gilead, Novartis, Abbvie, Eli Lilly, Pfizer, J&J, Mylan, UCB). Denis MONGIN, Romain Aymond, Rüdiger Müller and Delphine COURVOISIER have no conflicts of interest to disclose.

Funding

This analysis has been made possible by financial support of Eli Lilly (Suisse) SA to the Geneva University Hospitals (HUG). Also, the SCQM is financially supported by pharmaceutical

1
2
3 industries and donors, including Eli Lilly. A list of financial supporters can be found on
4
5 www.scqm.ch/sponsors .
6
7

8 **Data sharing statement**

9

10 Restrictions apply to the availability of these data. Data is owned by a third party, the Swiss
11
12 Clinical Quality Management in Rheumatic Diseases (SCQM) foundation. Data may be
13
14 obtained after approval and permission from this license holder (SCQM). Contact information
15
16 for data request: scqm@hin.ch
17
18
19

20 **Ethical Review and Regulatory Considerations**

21

22
23 This observational study has been approved by the Geneva ethical review boards (ERBs) as
24
25 required by local law (Project ID: 2019-00930 ; approval date 28 May 2019). Every participant
26
27 has signed an information and consent form at inclusion in the SCQM registry. Hence, this
28
29 study has been conducted in accordance with the ethical principles of the Declaration of
30
31 Helsinki and is consistent with Good Pharmacoepidemiology Practices (GPPs).
32
33
34

35 **Acknowledgements**

36

37
38 Almut Scherer, from the SCQM. Also, all of the participating physicians of the SCQM in
39
40 Rheumatic Diseases. A complete list of rheumatology offices and hospitals that are
41
42 contributing to the SCQM registries can be found on www.scqm.ch/institutions .
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

- 1 Emery P, Seto Y. Role of biologics in early arthritis. *Clin Exp Rheumatol* 2003;**21**:S191-194.
- 2 Lipsky PE, van der Heijde DMFM, St. Clair EW, *et al.* Infliximab and Methotrexate in the Treatment of Rheumatoid Arthritis. *New England Journal of Medicine* 2000;**343**:1594–602. doi:10.1056/NEJM200011303432202
- 3 Breedveld FC, Emery P, Keystone E, *et al.* Infliximab in active early rheumatoid arthritis. *Annals of the Rheumatic Diseases* 2004;**63**:149–55. doi:10.1136/ard.2003.013961
- 4 Klareskog L, van der Heijde D, de Jager JP, *et al.* Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *The Lancet* 2004;**363**:675–81. doi:10.1016/S0140-6736(04)15640-7
- 5 Weinblatt ME, Kremer JM, Bankhurst AD, *et al.* A Trial of Etanercept, a Recombinant Tumor Necrosis Factor Receptor:Fc Fusion Protein, in Patients with Rheumatoid Arthritis Receiving Methotrexate. *New England Journal of Medicine* 1999;**340**:253–9. doi:10.1056/NEJM199901283400401
- 6 Maini R, St Clair EW, Breedveld F, *et al.* Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. *The Lancet* 1999;**354**:1932–9. doi:10.1016/S0140-6736(99)05246-0
- 7 Breedveld FC, Weisman MH, Kavanaugh AF, *et al.* The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis & Rheumatism* 2006;**54**:26–37. doi:10.1002/art.21519
- 8 Moreland LW, Schiff MH, Baumgartner SW, *et al.* Etanercept Therapy in Rheumatoid Arthritis. *Annals of Internal Medicine* 1999;**130**:478–86. doi:10.7326/0003-4819-130-6-199903160-00004
- 9 Dixon WG, Carmona L, Finckh A, *et al.* EULAR points to consider when establishing, analysing and reporting safety data of biologics registers in rheumatology. *Annals of the Rheumatic Diseases* 2010;**69**:1596–602. doi:10.1136/ard.2009.125526
- 10 Wolfe F, Michaud K. The Hawthorne Effect, Sponsored Trials, and the Overestimation of Treatment Effectiveness. *The Journal of Rheumatology* 2010;**37**:2216–20. doi:10.3899/jrheum.100497
- 11 Wolfe F, Michaud K. The loss of health status in rheumatoid arthritis and the effect of biologic therapy: a longitudinal observational study. *Arthritis Res Ther* 2010;**12**:R35. doi:10.1186/ar2944

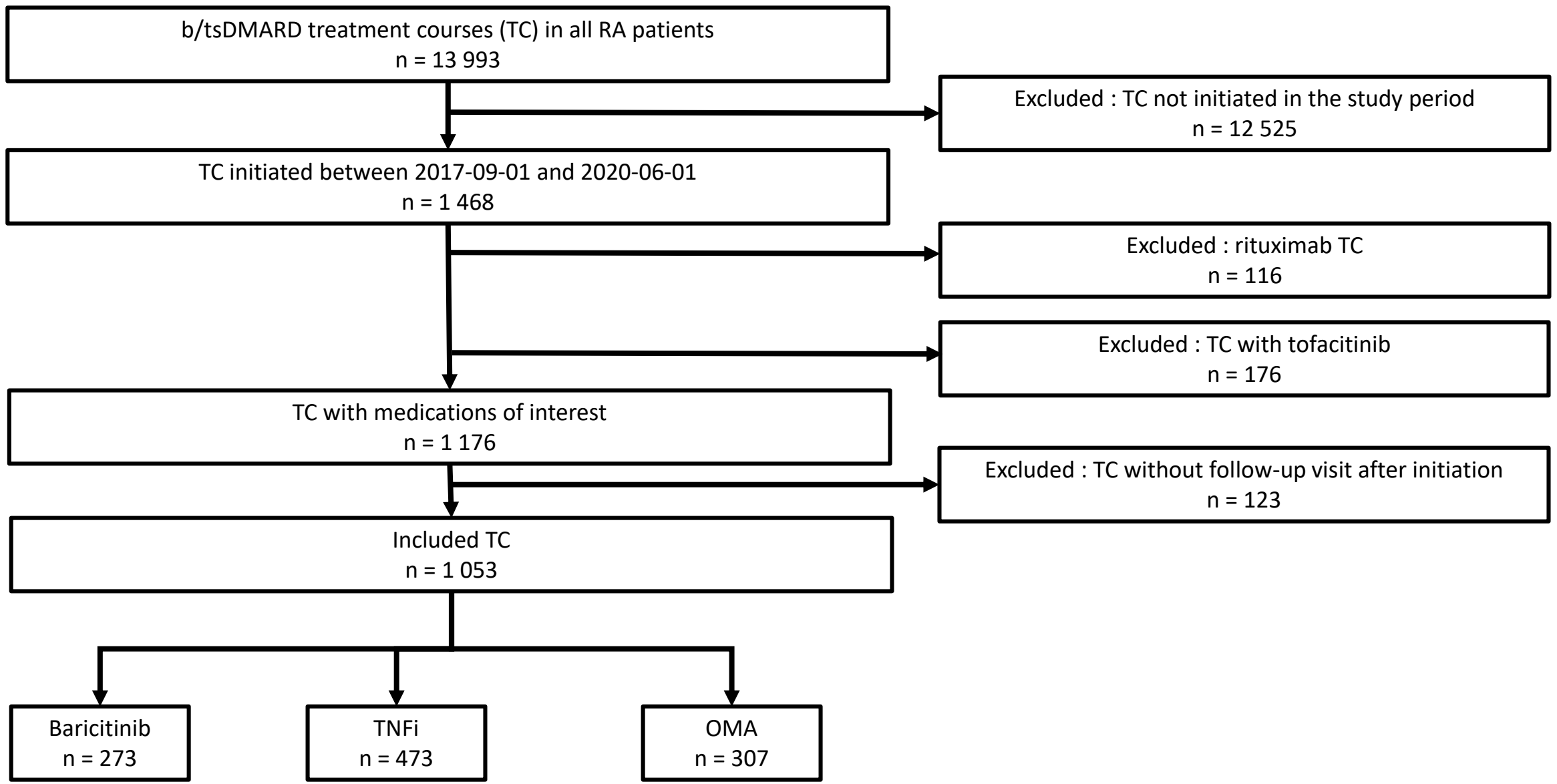
- 12 Østergaard M, Unkerskov J, Linde L, *et al.* Low remission rates but long drug survival in rheumatoid arthritis patients treated with infliximab or etanercept: results from the nationwide Danish DANBIO database. *Scandinavian Journal of Rheumatology* 2007;**36**:151–4. doi:10.1080/03009740601089267
- 13 Souto A, Maneiro JR, Gómez-Reino JJ. Rate of discontinuation and drug survival of biologic therapies in rheumatoid arthritis: a systematic review and meta-analysis of drug registries and health care databases. *Rheumatology* 2016;**55**:523–34. doi:10.1093/rheumatology/kev374
- 14 Taylor PC, Keystone EC, van der Heijde D, *et al.* Baricitinib versus Placebo or Adalimumab in Rheumatoid Arthritis. *New England Journal of Medicine* 2017;**376**:652–62. doi:10.1056/NEJMoa1608345
- 15 Fleischmann R, Schiff M, Heijde D van der, *et al.* Baricitinib, Methotrexate, or Combination in Patients With Rheumatoid Arthritis and No or Limited Prior Disease-Modifying Antirheumatic Drug Treatment. *Arthritis & Rheumatology* 2017;**69**:506–17. doi:10.1002/art.39953
- 16 Dougados M, van der Heijde D, Chen Y-C, *et al.* Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. *Ann Rheum Dis* 2017;**76**:88–95. doi:10.1136/annrheumdis-2016-210094
- 17 Genovese MC, Kremer J, Zamani O, *et al.* Baricitinib in Patients with Refractory Rheumatoid Arthritis. *N Engl J Med* 2016;**374**:1243–52. doi:10.1056/NEJMoa1507247
- 18 Smolen JS, Genovese MC, Takeuchi T, *et al.* Safety Profile of Baricitinib in Patients with Active Rheumatoid Arthritis with over 2 Years Median Time in Treatment. *J Rheumatol* 2019;**46**:7–18. doi:10.3899/jrheum.171361
- 19 Keystone EC, Taylor PC, Tanaka Y, *et al.* Patient-reported outcomes from a phase 3 study of baricitinib versus placebo or adalimumab in rheumatoid arthritis: secondary analyses from the RA-BEAM study. *Ann Rheum Dis* 2017;**76**:1853–61. doi:10.1136/annrheumdis-2017-211259
- 20 Taylor PC, Takeuchi T, Burmester GR, *et al.* Safety of baricitinib for the treatment of rheumatoid arthritis over a median of 4.6 and up to 9.3 years of treatment: final results from long-term extension study and integrated database. *Ann Rheum Dis* 2022;**81**:335–43. doi:10.1136/annrheumdis-2021-221276
- 21 Barbulescu A, Askling J, Chatzidionysiou K, *et al.* Effectiveness of baricitinib and tofacitinib compared with bDMARDs in RA: results from a cohort study using nationwide Swedish register data. *Rheumatology* 2022;:keac068. doi:10.1093/rheumatology/keac068
- 22 Pappas DA, John GS, Etzel CJ, *et al.* Comparative effectiveness of first-line tumour necrosis factor inhibitor versus non-tumour necrosis factor inhibitor biologics and targeted synthetic agents in patients with rheumatoid arthritis: results from a large US registry study. *Annals of the Rheumatic Diseases* 2021;**80**:96–102. doi:10.1136/annrheumdis-2020-217209

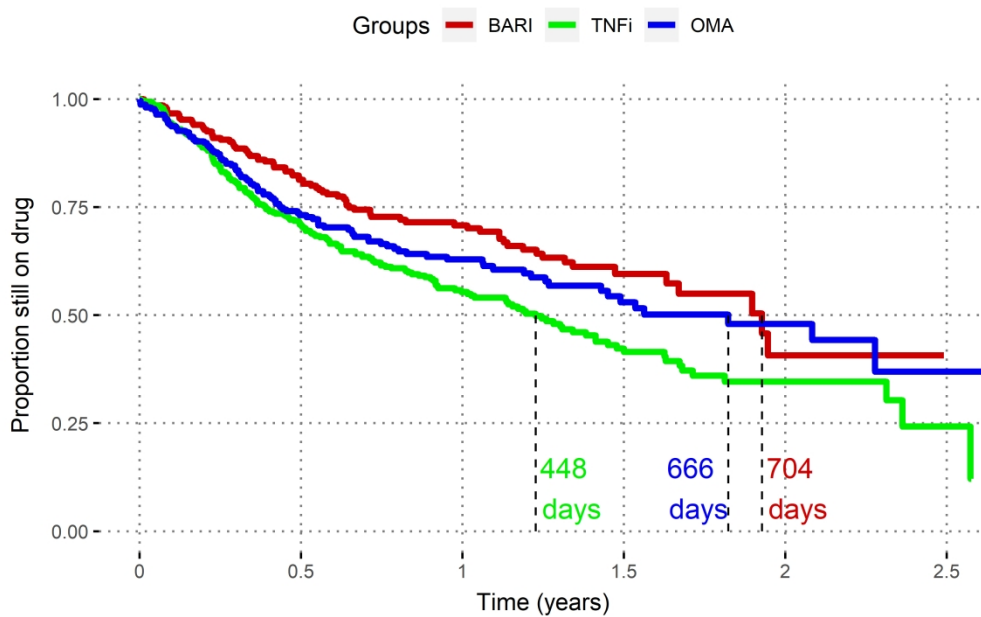
- 1
2
3 23 Alten R, Burmester GR, Cerinic MM, *et al.* Pos0666 a Multinational, Prospective,
4 Observational Study in Patients with Rheumatoid Arthritis Receiving Baricitinib, Targeted
5 Synthetic or Biologic Disease-Modifying Therapies: 12 Month Time to Discontinuation,
6 Effectiveness and Patient Reported Outcome Data from the European Cohort. *Annals of*
7 *the Rheumatic Diseases* 2022;**81**:606–7. doi:10.1136/annrheumdis-2022-eular.265
8
9
10 24 Pan SMD, Scherer A, Gabay C, *et al.* Differential drug retention between anti-TNF agents
11 and alternative biological agents after inadequate response to an anti-TNF agent in
12 rheumatoid arthritis patients. *Annals of the Rheumatic Diseases* 2012;**71**:997–9.
13 doi:10.1136/annrheumdis-2011-200882
14
15
16 25 Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease
17 Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis. ;:9.
18
19
20 26 Lauper K, Finckh A. Predictive factors of treatment persistence in rheumatoid arthritis.
21 *Joint Bone Spine* 2020;**87**:531–4. doi:10.1016/j.jbspin.2020.03.006
22
23 27 Mongin D, Lauper K, Finckh A, *et al.* Accounting for missing data caused by drug cessation
24 in observational comparative effectiveness research: a simulation study. *Ann Rheum Dis*
25 2022;:annrheumdis-2021-221477. doi:10.1136/annrheumdis-2021-221477
26
27
28 28 Mongin D, Lauper K, Turesson C, *et al.* Imputing missing data of function and disease
29 activity in rheumatoid arthritis registers: what is the best technique? *RMD Open*
30 2019;**5**:e000994. doi:10.1136/rmdopen-2019-000994
31
32
33 29 Buuren S van, Groothuis-Oudshoorn K. Mice : Multivariate Imputation by Chained
34 Equations in R. *J Stat Soft* 2011;**45**. doi:10.18637/jss.v045.i03
35
36
37 30 Finckh A, Tellenbach C, Herzog L, *et al.* Comparative effectiveness of antitumour necrosis
38 factor agents, biologics with an alternative mode of action and tofacitinib in an
39 observational cohort of patients with rheumatoid arthritis in Switzerland. *RMD Open*
40 2020;**6**:e001174. doi:10.1136/rmdopen-2020-001174
41
42
43 31 Lauper K, Iudici M, Mongin D, *et al.* Effectiveness of TNF-inhibitors, abatacept, IL6-
44 inhibitors and JAK-inhibitors in 31 846 patients with rheumatoid arthritis in 19 registers
45 from the ‘JAK-pot’ collaboration. *Annals of the Rheumatic Diseases* 2022;**81**:1358–66.
46 doi:10.1136/annrheumdis-2022-222586
47
48
49 32 Taylor PC, Lee YC, Fleischmann R, *et al.* Achieving Pain Control in Rheumatoid Arthritis
50 with Baricitinib or Adalimumab Plus Methotrexate: Results from the RA-BEAM Trial. *JCM*
51 2019;**8**:831. doi:10.3390/jcm8060831
52
53
54 33 Busch-Dienstfertig M, González-Rodríguez S. IL-4, JAK-STAT signaling, and pain. *JAKSTAT*
55 2013;**2**:e27638. doi:10.4161/jkst.27638
56
57
58 34 Álvaro-Gracia JM, García-Llorente JF, Valderrama M, *et al.* Update on the Safety Profile of
59 Tofacitinib in Rheumatoid Arthritis from Clinical Trials to Real-World Studies: A Narrative
60 Review. *Rheumatol Ther* 2021;**8**:17–40. doi:10.1007/s40744-020-00258-9

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Selection of eligible Treatment Courses (TC)





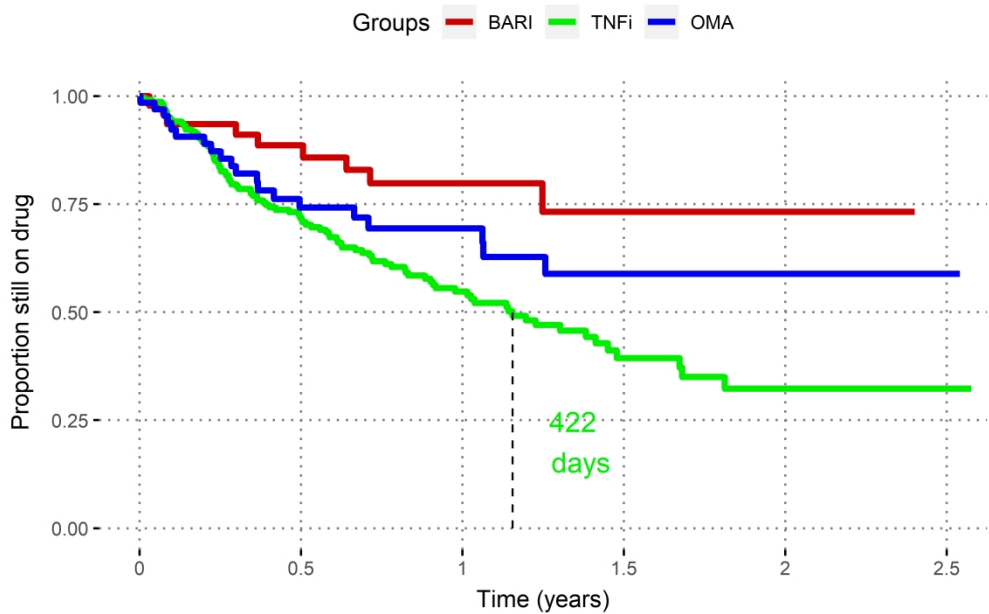
Number at risk

Groups	0	0.5	1	1.5	2	2.5
BARI	273	172	101	34	6	0
TNFi	473	247	129	51	19	3
OMA	307	159	94	37	18	3

Time (years)

451x386mm (197 x 197 DPI)

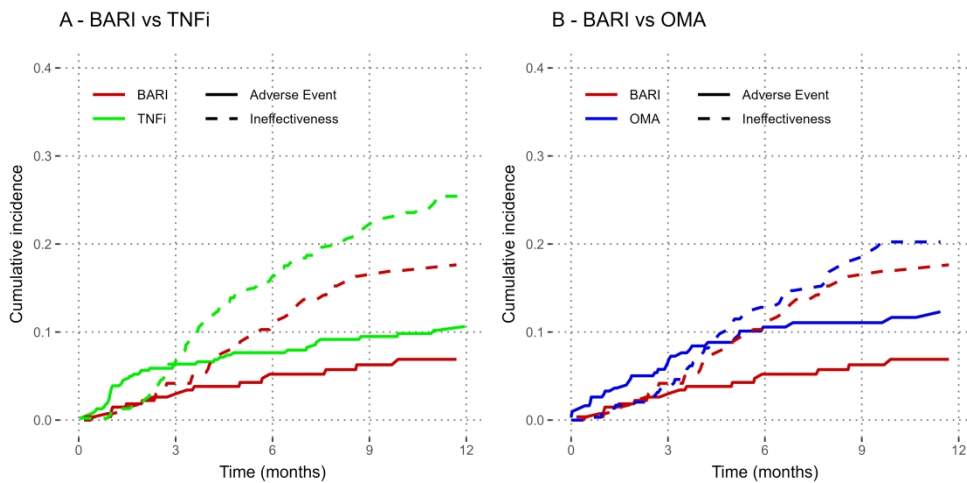
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Number at risk

Time (years)	0	0.5	1	1.5	2	2.5
BARI	46	31	17	5	1	0
TNFi	225	124	66	22	9	1
OMA	66	37	23	11	6	1

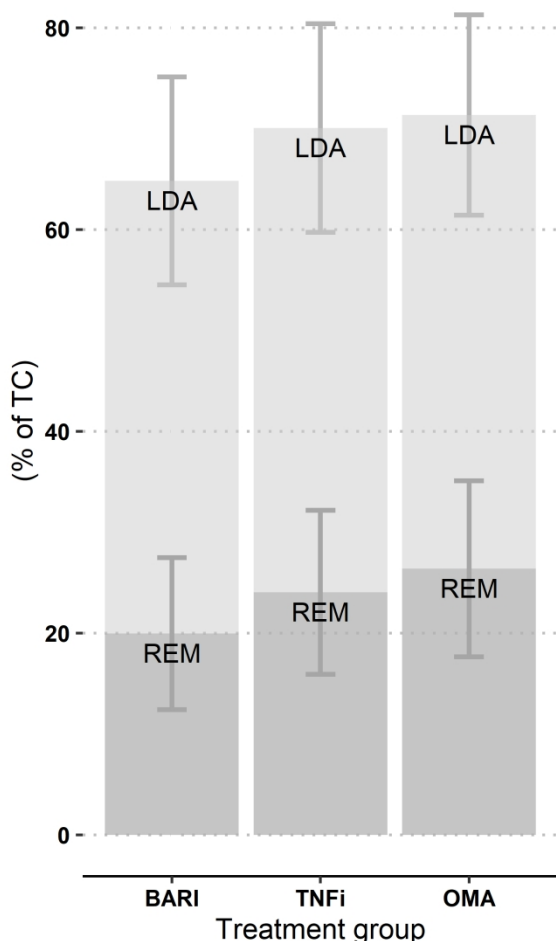
451x386mm (197 x 197 DPI)



644x322mm (197 x 197 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REM and LDA rates by type of treatment (CARRAC)



257x451mm (197 x 197 DPI)

SUPPLEMENTARY DATA

Notice on TC duration

Due to frequent changes in medication and short study period, it has to be underlined that the median duration of a TC approximates 200 days. The proportion of TC with follow-up data of at least one year is 37% for BARI, 27% for TNFi and 31% for OMA (Figure S1) - i.e. most TC were started less than 12 months before the date of data extraction.

Notice this % is different from the % of patient still under therapy that we estimate using Kaplan-Meier or Cox model. Indeed, the latter includes a censoring of the lost-to-follow-up patients, hence the denominator is different. As a consequence, this does not contradict the reported “median prescription survival timey”, for instance of 704 days for BARI TCs. The latter is the output estimated by the Kaplan-Meier model, taking censoring into account; it does not imply that actual observations in the dataset have this duration.

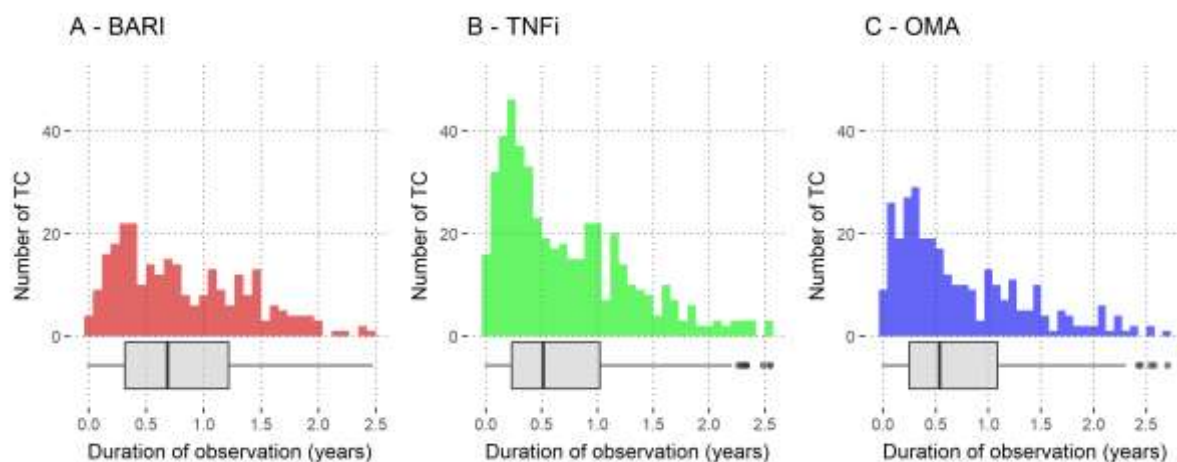


Figure S1: Distribution of the observation time for included TCs, per group, SCQM, 2017-2020.

Most of the treatment courses have an actual duration and/or follow-up period of less than one year. TC = Treatment Course. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs.

Variable definitions

Below we give additional detail about included covariates:

Age: age in years, at TC initiation. Continuous variable.

Gender: male or female. Categorical variable.

BMI: BMI at TC initiation. Continuous variable.

CDAI score: CDAI score at TC initiation. Continuous variable. If missing, imputed according to procedure described in methods section.

Disease duration: time interval between RA diagnosis date and TC initiation date. Continuous variable, expressed in years, but used in decades in models.

Smoking status: smoking status at TC initiation. Categorical variable (current-, former-, never-smoker).

Concomitant csDMARD: yes/no variable. A concomitant csDMARD was defined as csDMARD prescription ongoing for at least 40% of the duration of the TC. Otherwise, the TC was categorized as monotherapy. csDMARDs included: methotrexate, sulfasalazin, leflunomide, azathioprine and hydroxy-chloroquine, alone or in combination.

Concomitant glucocorticoid: Yes/no variable. Concomitant glucocorticoid usage was defined as having at least one active prescription of glucocorticoid, at any dose, at any timepoint of the TC.

Line of therapy: strictly speaking, this categorical variable is displaying: [*number of previous TC ever* + 1]. 4 or more has been grouped in the same category. Hence, it is considering all data of the SCQM registry, i.e. TCs initiated before our study period are also accounted for as previous therapies.

Seropositivity: yes/no variable. Seropositivity is defined as positivity for anti-citrullinated peptide antibodies and/or rheumatoid factor.

Time to all cause discontinuation

Cox model output

Table S1 contains the complete output of the two adjusted Cox models used in the main time-to-drug discontinuation analysis.

Table S1: Hazard ratio of drug discontinuation, Cox models, SCQM-RA registry, 2017-2020.						
	BARI vs TNFi			BARI vs OMA		
	Hazard ratio	95% CI	p	Hazard ratio	95% CI	p
TNFi (vs baricitinib)	1.76	1.32-2.35	<0.001	-	-	-
OMA (vs baricitinib)	-	-	-	1.27	0.93-1.72	0.13
Adjusting variables:						
Age (decades)	1.03	0.92-1.14	0.61	0.98	0.86-1.10	0.69
BMI	1.01	0.98-1.04	0.51	0.98	0.94-1.02	0.31
TC with csDMARD	0.84	0.66-1.09	0.19	1.22	0.90-1.67	0.20
Glucocorticoid usage	1.29	0.93-1.79	0.12	1.86	1.32-2.61	<0.001
CDAI score	1.40	1.26-1.56	<0.001	1.15	1.03-1.28	0.01
Disease duration (decades)	0.95	0.81-1.10	0.46	0.85	0.70-1.03	0.10
Current smoker (vs non-smoker)	1.20	0.86-1.68	0.28	1.09	0.73-1.64	0.66
Ever smoker (vs non-smoker)	1.10	0.79-1.52	0.57	1.38	0.95-2.00	0.09
2nd line therapy (vs 1st)	1.11	0.80-1.53	0.52	1.37	0.81-2.33	0.24
3rd line therapy (vs 1st)	0.98	0.64-1.51	0.93	1.56	0.92-2.64	0.10
4th or later line (vs 1st)	1.06	0.75-1.51	0.73	1.57	0.93-2.63	0.09
Female gender	1.05	0.78-1.42	0.74	1.16	0.81-1.67	0.41
Seropositivity	0.77	0.59-1.01	0.055	0.94	0.67-1.31	0.71

Table S1: BARI = baricitinib. TNFi = TNF inhibitors. OMA = Other Mode of Action drugs. CI = Confidence Interval. BMI = Body Mass Index. CDAI = Clinical Disease Activity Index.

Figure S2 below gives the exact same information as Table S1:

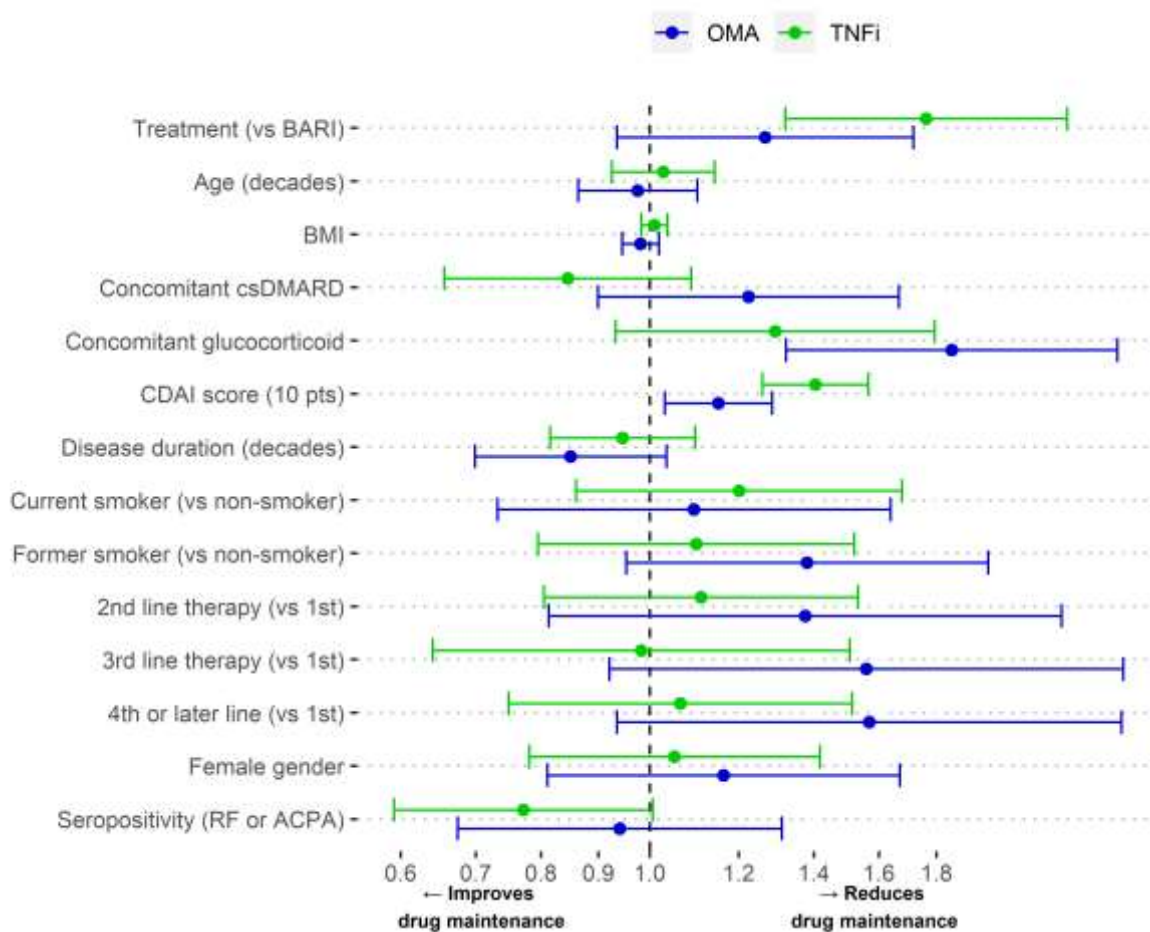


Figure S2: Hazard ratio of drug discontinuation (95% CI).

BARI = Baricitinib. TNFi = Tumor Necrosis Factors Inhibitors. OMA = Other mode of Action bDMARDs. BMI = Body Mass Index. CDAI = Clinical Disease Activity Index. RF = Rheumatoid Factor. ACPA = Anti-citrullinated Peptides Antibodies.

The corresponding cox-adjusted drug-survival curves are provided below:

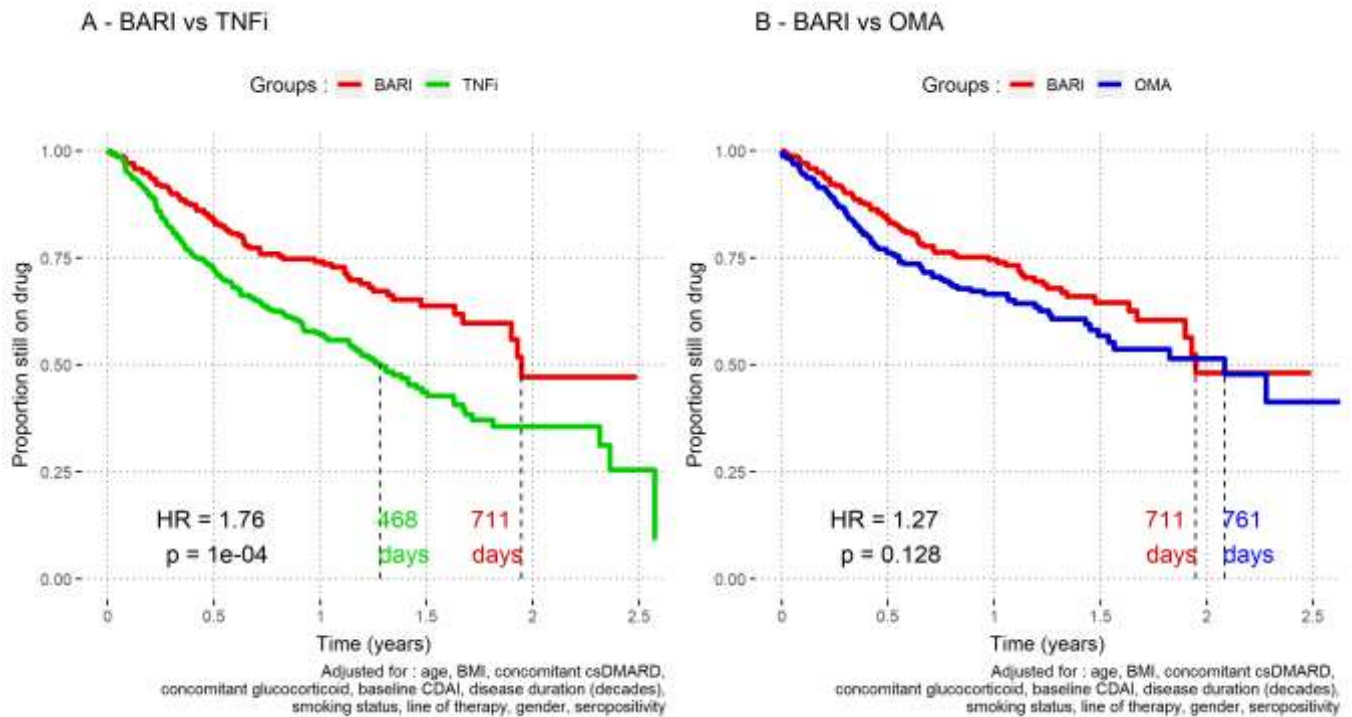


Figure S3: Multivariable Cox model of drug discontinuation by type of treatment, SCQM, 2017-2020.

These curves are merely the visualisation of Cox models presented in Table S1 and Figure S2.

BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action bDMARDs

Models are adjusted for : age, BMI, concomitant csDMARD, concomitant glucocorticoid, baseline CDAI, disease duration, smoking status, line of therapy, gender, seropositivity.

Sensitivity analysis using AIPTW

As a sensitivity analysis, the main time to drug discontinuation was also performed using “augmented inverse probability of treatment weighting” (AIPTW), including the same covariates. In other words, we combined a propensity score using a logistic regression model and an inverse probability weighted Cox regression. We used the *RiskRegression* package in R, to obtain risk ratios.

Figure S4 represents the absolute risk of treatment discontinuation, for all included timepoints. At one year, the adjusted discontinuation risk in BARI was 19 % lower than in TNFi group ($p < 0.001$) (Figure S4 A), with a risk ratio of 1.76 (95% CI [1.19 – 2.34]; $p = 0.009$). Similarly, at one year, the adjusted treatment discontinuation risk in BARI was 8 % lower than in the OMA group ($p = 0.06$) (Figure S4 B), with a risk ratio of 1.28 (95% [0.91 – 1.65]; $p = 0.14$).

Overall, this sensitivity analysis confirms the findings reported in the main body of the article.

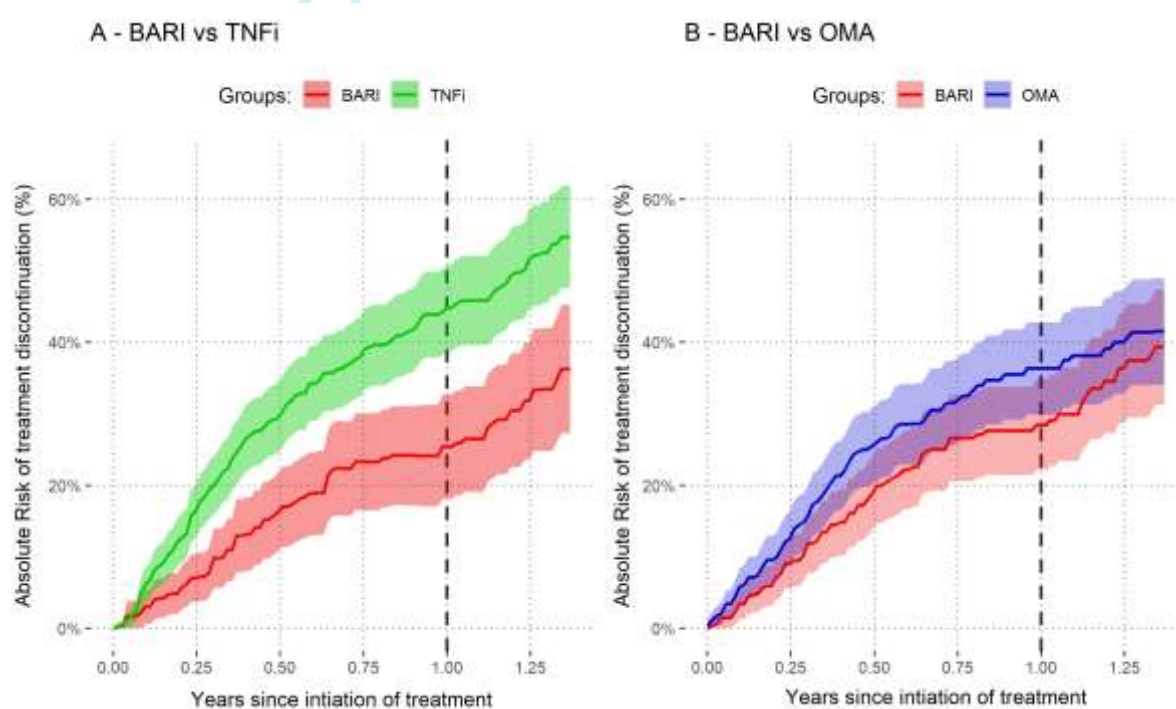


Figure S4: Absolute risk of treatment discontinuation by type of treatment (AIPTW), SCQM, 2017-2020

BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action bDMARDs.

AIPTW = Augmented Inverse Probability of Treatment Weighting. Adjusted for : age, bmi, concomitant csDMARDs, prednisone usage, baseline CDAI, disease duration, smoking status, line of therapy, gender, seropositivity.

Time to all-cause-discontinuation in b/tsDMARD-naïve patients

Table S2: Baseline characteristics of study population, b/tsDMARD-naïve patients, SCQM-RA registry, 2017-2020.

Patient-Variables	BARI (n = 46)		TNFi (n = 225)		OMA (n = 66)		p values
	n %	Miss.	n %	Miss.	n %	Miss.	
Female	70 %	0	71 %	1	73 %	1	0.88
Age (years)	57 (15)	0	52 (14)	1	57 (16)	1	<0.01
Disease duration (years)	6 (6)	1	5 (7)	13	7 (9)	2	0.24
CDAI baseline (raw data)	16 (8)	31	18 (10)	135	18 (14)	42	0.77
CDAI baseline (imputed)	12 (7)	0	14 (9)	0	14 (10)	0	0.61
Obesity (BMI > 30)	11 %	13	13 %	58	5 %	27	0.28
Smoking							
Current	28 %	9	18 %	42	14 %	13	0.18
Former	26 %		24 %		21 %		
Never	26 %		39 %		46 %		
Seropositive (ACPA or RF)	80 %	1	69 %	5	76 %	2	0.20
TC variables							
Dose of BARI (4mg)	83 %	0	-	-	-	-	-
TC duration > 12-months	37 %	0	29 %	0	35 %	0	0.48
Concomitant csDMARD	41 %	0	66 %	0	50 %	0	<0.01
Line of Therapy							
1 st (= bio-naïve)							
2 nd	100 %	0	100 %	0	100 %	0	-
3 rd							
4 th or later							
Previous tsDMARD use (non-BARI)	0 %	0	0 %	0	0 %	0	-
Concomitant glucocorticoid (at any time)	13 %	0	20 %	0	17 %	0	0.50
Mean dose of concomitant glucocorticoid (mg)	1 (4)	0	3 (6)	0	2 (6)	0	0.50

Table S2: BARI = baricitinib. TC = Treatment Courses. CDAI = Clinical Disease Activity Index. TNFi = Tumor Necrosis Factor Inhibitors. OMA = bDMARDs with Other Mode of Action. tsDMARD = targeted synthetic DMARDs. ACPA = Anti Citrullinated Peptide Antibodies. RF = Rheumatoid Factor. Miss. = number of missing values. p-value are computed with either Chi² or ANOVA.

Response rates – raw CDAI data

Figure S6 shows the crude available values for CDAI scores, by type of treatment and time.

Only a minority of CDAI scores were assessed at 0- or 12-month timepoints of TCs (i.e. 680/1053 = 65% were missing for baseline value, and 908/1053 = 86% were missing for exact 12-month value). Future research would certainly benefit having CDAI scores assessed at regular and homogenous time-intervals, based on the initiation date of biological therapies.

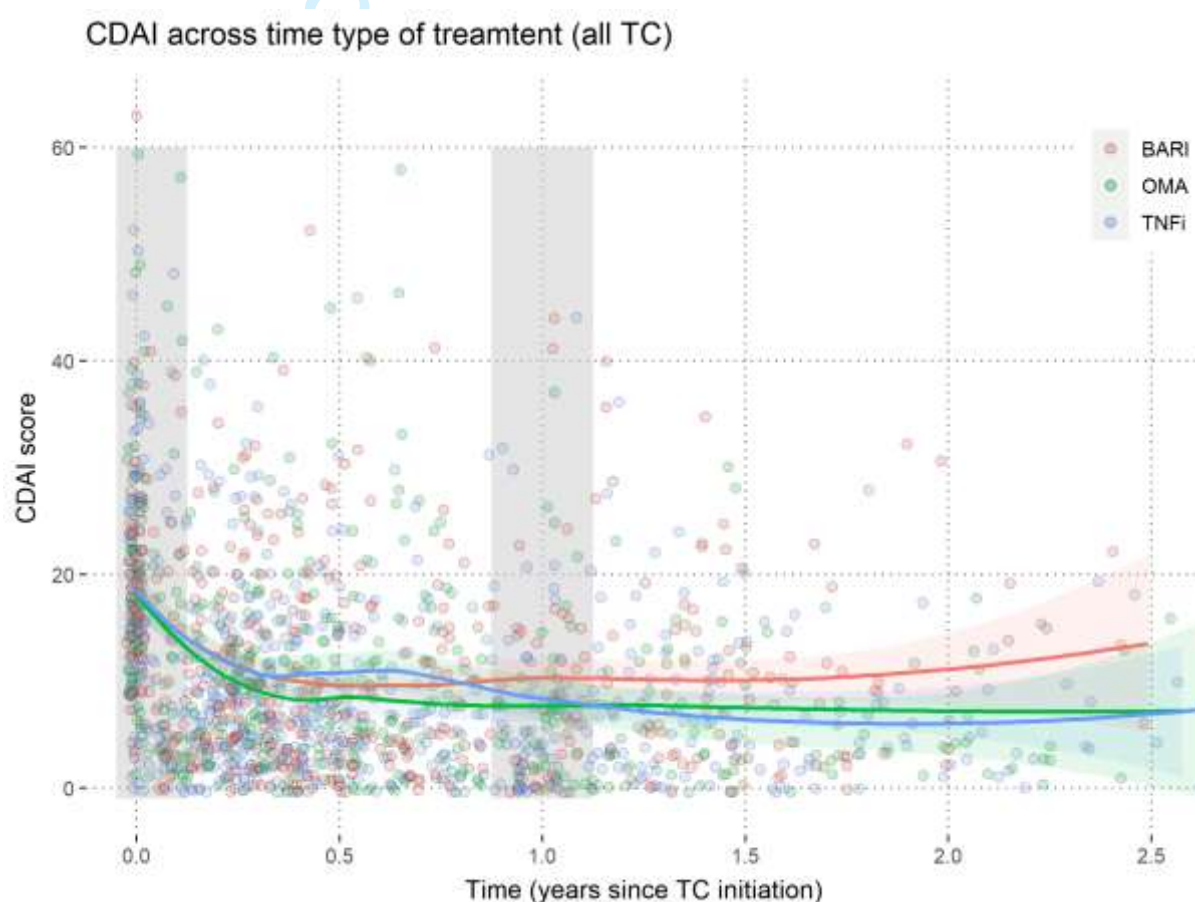


Figure S5: CDAI across time by type of treatment, raw data, SCQM, 2017-2020.

Only a minority of CDAI score were obtained sharp at 0 or 12 months of TCs. CDAI = Clinical Disease Activity Index. TC = Treatment course. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action b/tsDMARDs.

Study size

Based on estimates from similar analyses with tofacitinib (TOFA) performed in this registry, we calculated the number of patients that would be needed to detect a significant decrease in time to all cause-discontinuation of treatment (hazard ratio) between treatment groups using the method described by Schoenfeld and Richter. We assumed a statistical power of 80%, a type I error probability of 0.05, a median BARI retention of 30 months, the inclusion of 3 patients on TNFi for every patient on BARI, an accrual time of 2 years, and additional follow-up of 6 months. We display below the sample size for the BARI group for a range of possible effect sizes (“hazard ratio” between 1.1 and 1.8).

If the true hazard ratio is similar to the one found with TOFA compared to TNFi after a single TNFi failure (HR:1.68) 14, we will need to study 149 patients on BARI and 447 patients on TNFi to be able to reject the null hypothesis that the experimental and control curves are equal with probability (power) of 80%. Pragmatically, we propose to start the analysis of the data only once at least 200 patients on BARI have been included and followed for an average of at least 18 months.

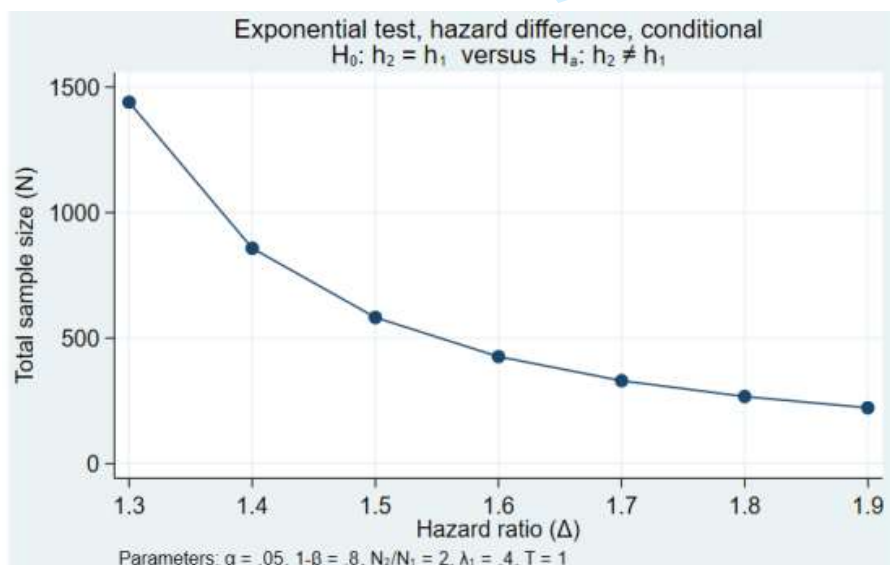


Figure 1: Estimated total sample size for two-sample comparison of survivor functions

1 - SURVIVAL ANALYSIS

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)
library(gridExtra)
library(grid)
library(cmprsk)
library(mstate)
library(cobalt)

rm(list = ls()) # To select all loaded objects and delete them
setwd(dirname(rstudioapi::getActiveDocumentContext())$path) # setting
up working directory in the location of the .Rmd file

load("./1_datamanaged_files/datamanaged.Rdata") # loading data.managed
data

Loading fonctions

# home-made function to force writing with two decimals
formattable = function(nbr){return(formatC(nbr,format = "f",digits =
nombreadpresvirgule))}
nombreadpresvirgule <- 2

# Home-made Fonction to write the p value (by denis)
writepvalue = function(pvalue) {
  if (is.na(pvalue)) {result <- NA} else {
    if(pvalue < 0.001) {
```

```

1
2
3     result <- "<0.001"
4   } else if (pvalue <0.01) {
5     result <- formatC( pvalue ,format = "f",digits = 3)
6   }
7   else {
8     (result <- formatC( pvalue ,format = "f",digits = 2) )
9     i = 1
10    while(result == 0.05) {
11      result <- formatC( pvalue ,format = "f",digits = 2 + i)
12      i = i + 1
13    }
14  }
15  return(result)
16 }
17 }
18 options(scipen = 999)
19
20
21

```

Mini Exploration

```

22 uniqueN(BARI_DATA[]$patient_id) # number of patients (< than number of
23 TC)
24 uniqueN(BARI_DATA[]$TC_id) # unumber of TC
25
26

```

```

27 plot <- qplot(x = BARI_DATA[]$time_on_drug)+
28   geom_vline(xintercept = 365, size = 1.2, alpha = 0.5)+
29   geom_text(aes(x = 365 + 40, label="1 Year", y=20), colour="white",
30   angle=0)+
31
32   geom_vline(xintercept = 2*365, size = 1.2, alpha = 0.5)+
33   geom_text(aes(x = 2*365 + 40, label="2 Year", y=20), colour="black",
34   angle=0)+
35
36   geom_vline(xintercept = mean(BARI_DATA$time_on_drug), color = "red",
37   size = 1.2)+
38   geom_text(aes(x = mean(BARI_DATA$time_on_drug) + 40, label="Mean",
39   y=20), colour="red", angle=0)+
40
41   geom_vline(xintercept = median(BARI_DATA$time_on_drug), color =
42   "green", size = 1.2)+
43   geom_text(aes(x = median(BARI_DATA$time_on_drug) - 40,
44   label="Median", y=20), colour="green", angle=0)+
45
46   labs(x = "Duration of TC (days)", y = "Number of TC", title =
47   "Repatriation of the duration of included TC (all groups)")+
48
49   theme_pubclean()
50
51
52
53 plot
54
55 mean(BARI_DATA$time_on_drug)
56
57
58
59
60

```

```

1
2
3 median(BARI_DATA$time_on_drug)
4
5 # Nb : Research protocol said we wanted a follow-up duration of
6 "average of 18 months"
7 mean(BARI_DATA[cohort == "BARI"]$time_on_drug)/30
8 mean(BARI_DATA[cohort == "TNFi"]$time_on_drug)/30
9 mean(BARI_DATA[cohort == "OMA"]$time_on_drug)/30 # looks more like 9
10 months..
11
12
13 # ok, 24-month follow-up will be complicated
14 uniqueN(BARI_DATA[time_on_drug > 2*365, TC_id]) # number if TC with
15 duration > 24 months
16
17
18

```

1. [0] Table 1 BARI vs TNFi et OMA bDMARDs

Common table with all the data

Showing NA to have complete counts and accurate % in each category

```

24 BARI_DATA[,time_on_drugDiff0 := as.numeric(time_on_drug > 0)] # time
25 in drug < 0

```

```

26 BARI_DATA[,time_on_drug365 := as.numeric(time_on_drug > 365.25)]
27

```

```

28
29 myVars2 <- c("gender", "age_base", "disease_duration_base_years",
30 "CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
31 "seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
32 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
33 "PREDNISON_STEROID_dose", "dose", "initiation_year",
34 "time_on_drug", "HAQ_score_base")
35

```

```

36 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
37 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
38 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
39 "seropositivity_base")
40

```

```

41 nonnormalVars2 <- c()

```

```

42 tab2 <- CreateTableOne(vars = myVars2, data = BARI_DATA, factorVars =
43 catVars2, strata = "cohort", test = F, includeNA = T)

```

```

44 tablexp2 <- print(tab2, nonnormal= nonnormalVars2, catDigits = 1,
45 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
46

```

saving table 1 NA

```

49 write.xlsx(tablexp2, file =
50 "./3_clean_output/BARI_3_groups_table1_NA.xlsx")
51

```

Without NA to obtain adequate p values

```

54 BARI_DATA[,time_on_drugDiff0 := as.numeric(time_on_drug > 0)] # time
55 in drug < 0
56
57
58
59

```

```
1
2
3 BARI_DATA[,time_on_drug365 := as.numeric(time_on_drug > 365.25)]
4
5 myVars2 <- c("gender", "age_base","disease_duration_base_years",
6 "CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
7 "seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
8 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
9 "PREDNISON_STEROID_dose", "dose", "initiation_year",
10 "time_on_drug","HAQ_score_base")
11
12
13 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
14 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
15 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
16 "seropositivity_base")
17
18 nonnormalVars2 <- c()
19 tab2 <- CreateTableOne(vars = myVars2, data = BARI_DATA, factorVars =
20 catVars2, strata = "cohort", test = T, includeNA = F)
21 tablexp2 <- print(tab2, nonnormal= nonnormalVars2, catDigits = 1,
22 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
23
```

24 Saving table 1

```
25
26 write.xlsx(tablexp2, file =
27 "./3_clean_output/BARI_3_groups_table1.xlsx")
28
```

29 Other various computations for Table 1

```
30 uniqueN(BARI_DATA$patient_id)
31
32
33 mean(BARI_DATA[, time_on_drug])
34
35 median(BARI_DATA[cohort == "Bari" , time_on_drug])
36 median(BARI_DATA[cohort == "OMA" , time_on_drug])
37 median(BARI_DATA[cohort == "TNFi" , time_on_drug])
38
39 median(BARI_DATA[cohort == "OMA"]$time_on_drug)
40
41
42 mean(BARI_DATA[,disease_duration_base_years], na.rm = T)
43
44 table(is.na(BARI_DATA$CDAI0_raw), BARI_DATA$cohort) # number of
45 missing CDAI0_raw...
46 table(is.na(BARI_DATA$CDAI0), BARI_DATA$cohort) # number of missing
47 CDAI0... (after imputation)
48
49 table(is.na(BARI_DATA$CDAI12_raw), BARI_DATA$cohort) # number of
50 missing CDAI12_raw...
51 table(is.na(BARI_DATA$CDAI12), BARI_DATA$cohort) # number of missing
52 CDAI12 after imputation
53
54
55 hist(BARI_DATA$CDAI0_raw)
56
57
58
59
```

```
1
2
3 hist(BARI_DATA$CDAI0)
4
5 summary(BARI_DATA[cohort=="BARI", c("gender",
6 "age_base", "disease_duration_base_years", "CDAI0_raw", "CDAI0",
7 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
8 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
9 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
10 "initiation_year", "time_on_drug", "HAQ_score_base")]) # to see NA
11 values for all variables
12
13
14 summary(BARI_DATA[cohort=="TNFi", c("gender",
15 "age_base", "disease_duration_base_years", "CDAI0_raw", "CDAI0",
16 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
17 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
18 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
19 "initiation_year", "time_on_drug", "HAQ_score_base")]) # to see NA
20 values for all variables
21
22
23 summary(BARI_DATA[cohort=="OMA", c("gender",
24 "age_base", "disease_duration_base_years", "CDAI0_raw", "CDAI0",
25 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
26 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
27 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
28 "initiation_year", "time_on_drug", "HAQ_score_base")]) # to see NA
29 values for all variables
30
31 table(is.na(BARI_DATA$disease_duration_base_years), BARI_DATA$cohort)
32 # number of missing disease duration...
33
34 table(is.na(BARI_DATA$age_base), BARI_DATA$cohort) # number of steroid
35 doses missing
36
37
38 table(is.na(BARI_DATA$PREDNISON_STEROID_dose), BARI_DATA$cohort) #
39 number of missing baseline steroids
40
```

Imputation using MICE -- BARI vs TNFi et OMA bDMARDs --

Common imputation step with all data

```
44
45 BARI <- BARI_DATA[,c("TC_id", "patient_id", "stop_DMARD",
46 "stop_reasons", "age_base", "concomitant_csDMARD",
47 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
48 "CDAI0", "disease_duration_base_years", "time_on_drug", "bmi_base",
49 "smoker_base", "line_of_therapy", "obese_base", "gender", "cohort",
50 "adverse_event_reported", "seropositivity_base")] # choose variables
51 of interest
52
```

```
53
54 BARI$smoker_base <- as.factor(BARI$smoker_base) # put labels as factor
55 BARI$line_of_therapy <- as.factor(BARI$line_of_therapy)
56 BARI$gender <- as.factor(BARI$gender)
57
58
59
```

```
1
2
3 BARI$concomitant_csDMARD <- as.factor(BARI$concomitant_csDMARD)
4 BARI$PREDNISON_STEROID <- as.factor(BARI$PREDNISON_STEROID)
5 BARI$cohort <- as.factor(BARI$cohort)
6
7 # Imputation
8
9
10 if(!file.exists("./2_cached_files/imputed_data")){ # to avoid re-
11 computing if already done
12   imputed_data <- mice(BARI, m=50, method="pmm", maxit=25, seed=500)
13   save(imputed_data, file = "./2_cached_files/imputed_data")
14 } else {
15   load("./2_cached_files/imputed_data")
16 }
17
18 # Subsettings
19
20 BARI1 <- BARI[cohort %in% c("BARI", "TNFi")] # creating subset for
21 BARI vs TNFi comparaisn
22 BARI1[,cohort := as.factor(as.character(cohort))]
23
24
25 imputed_data1 <- complete(imputed_data,"long", include=T) # to put in
26 long format and categorize variables
27 imputed_data1 <- imputed_data1[imputed_data1$cohort %in% c("BARI",
28 "TNFi"),] # to keep only BARI and TNFi rows
29 imputed_data1$cohort <- as.factor(as.character(imputed_data1$cohort))
30 imputed_data1 <- as.mids(imputed_data1) # re concateneting in previous
31 format, to use fit.mult.impute
32
33
34
35 BARI2 <- BARI[cohort %in% c("BARI", "OMA")] # creating subset for BARI
36 vs OMA comparaisn
37 BARI2[,cohort := as.factor(as.character(cohort))]
38
39
40 imputed_data2 <- complete(imputed_data,"long", include=T) # to put in
41 long format and categorize variables
42 imputed_data2 <- imputed_data2[imputed_data2$cohort %in% c("BARI",
43 "OMA"),] # to keep only BARI and OMA rows
44 imputed_data2$cohort <- as.factor(as.character(imputed_data2$cohort))
45 imputed_data2 <- as.mids(imputed_data2) # re concateneting in previous
46 format, to use fit.mult.impute
47
48
```

1. [1] SURVIVAL ANALYSIS (drug discontinuation)

Exploration

```
51 table(BARI_DATA$cohort, BARI_DATA$stop_DMARD)
52 table(BARI_DATA$cohort, BARI_DATA$stop_reasons)
53
54
55
56
57
58
59
```

Checking adequacy of COX models --

For BARI vs TNFi

```
1
2
3
4
5
6
7 # categorization for linearity checking
8 test1 <- complete(imputed_data1,"long", include=T) # to put in long
9 format and categorize variables
10
11 test1$agecat <- cut(test1$age_base, 4)
12 test1$bmicat <- cut(test1$bmi_base, 4)
13 test1$cdaicat <- cut(test1$CDAI0, 4)
14 test1$duracat <- cut(test1$disease_duration_base_years, 4)
15
16
17 test1 <- as.mids(test1, .imp=1, .id=2) # re concatenating in previous
18 format, to use fit.mult.impute
19
20 # linearity checking
21
22 BARI1.adj.mi.test <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
23 as.factor(cohort)+
24
25 I(agecat)+
26 I(bmicat)+
27 TC_with_csDMARD+
28 PREDNISON_STEROID+
29 I(cdaicat)+
30 I(duracat)+
31 C(smoker_base, base=3)+
32 line_of_therapy+
33 gender+
34 seropositivity_base+
35 cluster(patient_id),
36 fitter = coxph, xtrans = test1,
37 data = BARI1)
38
39 summary(BARI1.adj.mi.test)
40 rm(BARI1.adj.mi.test)
41
42
43 # Log-linearity of coefficients ?
44
45 # Coefs age are between 0.15 and 0.25, let's assume it's ok
46 # Hum bmi coefs are not so log-linear, rather close to 0
47 # For CDAI also
48 # Looks ok for disease_duration_base_years.
49
50
51 # --> Let's keep all variable in the continuous format
52
53 # Proportionality test of hazards on raw data
54
55 test1ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
```

```
1
2
3 as.factor(cohort)+
4             cluster(patient_id),
5             data= BARI1)
6 cox.zph(test1ph) # it's ok
7
8 # Hazard proportionality test on imputed data sets
9
10 test1 <- complete(test1,"long",include=T) # To reset the charges to
11 long format
12 test1 <- test1[test1$.imp==1 | test1$.imp==2 | test1$.imp==3 |
13 test1$.imp==4 | test1$.imp==5 ,] # To select only 5 data sets
14
15
16 test1ph.adj.mi <- coxph(Surv(time = time_on_drug, event = stop_DMARD)
17 ~ as.factor(cohort)+
18             I(age_base/10)+
19             bmi_base+
20             TC_with_csDMARD+
21             PREDNISON_STEROID+
22             CDAI0+
23             I(disease_duration_base_years/10)+
24             C(smoker_base, base=3)+
25             line_of_therapy+
26             gender+
27             seropositivity_base+
28             cluster(patient_id),
29             data = test1)
30
31
32 cox.zph(test1ph.adj.mi)
33
34 schonfeldall <- cox.zph(test1ph.adj.mi) # Test cox.zph may not be ok,
35 but it's because of the multiple imputation (often the case with a lot
36 of data)
37 for (i in 1:11){
38     plot(schonfeldall[i]) # so we should go to a visual testing --> ok
39 }
40
41
42 rm(schonfeldall, test1ph.adj.mi, test1)
43
44 For BARI vs OMA
45
46 # categorization for linearity checking
47 test2 <- complete(imputed_data2,"long", include=T) # to put in long
48 format and categorize variables
49
50 test2$agecat <- cut(test2$age_base, 4)
51 test2$bmicat <- cut(test2$bmi_base, 4)
52 test2$cdaicat <- cut(test2$CDAI0, 4)
53 test2$duracat <- cut(test2$disease_duration_base_years, 4)
54
55
56 test2 <- as.mids(test2, .imp=1, .id=2) # re concatenating in previous
57
58
59
60
```



```
1
2
3     format, to use fit.mult.impute
4
5     # linearity checking
6
7     BARI2.adj.mi.test <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
8     as.factor(cohort)+
9
10                                I(agecat)+
11                                I(bmicat)+
12                                TC_with_csDMARD+
13                                PREDNISON_STEROID+
14                                I(cdaicat)+
15                                I(duracat)+
16                                C(smoker_base, base=3)+
17                                line_of_therapy+
18                                gender+
19                                seropositivity_base+
20                                cluster(patient_id),
21                                fitter = coxph, xtrans = test2,
22
23     data = BARI2)
24
25     summary(BARI2.adj.mi.test)
26     rm(BARI2.adj.mi.test)
27
28     # Log-linearity of coefficients ?
29
30     # Coefs age are around -0.4, let's assume it's ok
31     # Hum bmi coefs are discussable
32     # For CDAI it's ok
33     # More or less ok for disease_duration_base_years.
34
35     # --> Let's keep all variable in the continuous format
36
37     # Proportionality test of hazards on raw data
38
39
40     test2ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
41     as.factor(cohort)+
42                                cluster(patient_id),
43                                data= BARI2)
44     cox.zph(test2ph) # it's ok
45
46
47
48     # Hazard proportionality test on imputed data sets
49
50     test2 <- complete(test2,"long",include=T) # To put imputed data in
51     ling format
52     test2 <- test2[test2$.imp==1 | test2$.imp==2 | test2$.imp==3 |
53     test2$.imp==4 | test2$.imp==5 ,] # To select only 5 datasets
54
55     test2ph.adj.mi <- coxph(Surv(time = time_on_drug, event = stop_DMARD)
```

```

1
2
3 ~ as.factor(cohort)+
4           I(age_base/10)+
5           bmi_base+
6           TC_with_csDMARD+
7           PREDNISON_STEROID+
8           CDAI0+
9           I(disease_duration_base_years/10)+
10          C(smoker_base, base=3)+
11          line_of_therapy+
12          gender+
13          seropositivity_base+
14          cluster(patient_id),
15          data=test2)
16
17
18 cox.zph(test2ph.adj.mi)
19
20 schonfeldall <- cox.zph(test2ph.adj.mi) # Test cox.zph may not be ok,
21 but it's because of the multiple imputation (often the case with a lot
22 of data)
23 for (i in 1:11){
24   plot(schonfeldall[i]) # so we should go to a visual testing --> ok
25 }
26
27
28 rm(schonfeldall, test2ph.adj.mi, test2)
29

```

BARI vs TNFi --

COX model

Final Cox Model

```

36 BARI1.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
37 cohort +
38           I(age_base/10)+
39           bmi_base+
40           TC_with_csDMARD+
41           PREDNISON_STEROID+
42           I(CDAI0/10)+
43           I(disease_duration_base_years/10)+
44           C(smoker_base, base=3)+
45           line_of_therapy+
46           gender+
47           seropositivity_base+
48           cluster(patient_id),
49           fitter = coxph, xtrans = imputed_data1,
50
51 data = BARI1)
52
53 summary(BARI1.adj.mi)
54
55
56
57
58
59
60

```

Creation of HR table and p-values

```

1
2
3
4 ploufrows <- names(BARI1.adj.mi$coefficients)
5 ploufcols <- c("HR","95%CI","p")
6 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
7 length(ploufcols))
8 rownames(coxtable) <- ploufrows
9 colnames(coxtable) <- ploufcols
10 plouf <- summary(BARI1.adj.mi)
11
12
13 for(row in ploufrows)
14 {
15   coxtable[row,"HR"] <-
16   formattable(plouf$coefficients[row,"exp(coef)"])
17   coxtable[row,"95%CI"] <-
18   paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",
19   formattable(plouf$conf.int[row,"upper .95"]))
20   coxtable[row,"p"] <- writepvalue(plouf$coefficients[row,"Pr(>|
21 z|)"])
22 }
23
24
25 write.xlsx(coxtable, file="./3_clean_output/BARI vs TNFi HR.xlsx") #
26 saving excel file
27

```

Forest plot

```

28
29 meanall <- summary(BARI1.adj.mi)$coefficients[1:14,"exp(coef)"]
30 lowerall <- summary(BARI1.adj.mi)$conf.int[1:14,"lower .95"]
31 upperall <- summary(BARI1.adj.mi)$conf.int[1:14,"upper .95"]
32 textall <- c("TNFi (vs BARI)", "Age (decades)", "BMI", "Concomitant
33 csDMARD", "Concomitant glucocorticoid", "CDAI score (10 pts)",
34 "Disease duration (decades)", "Current smoker (vs non-smoker)",
35 "Former smoker (vs non-smoker)", "2nd line therapy (vs 1st)", "3rd
36 line therapy (vs 1st)", "4th or later line (vs 1st)", "Female gender",
37 "Seropositivity (RF or ACPA)")
38
39
40 dfall <- data.frame(textall, meanall, lowerall, upperall)
41 dfall$textall <- factor(dfall$textall,
42 levels = textall)
43
44 HR_plot_1 <- ggplot(data=dfall, aes(x=textall, y= meanall, ymin =
45 lowerall, ymax = upperall))+
46
47   geom_pointrange(size=0.5)+
48   geom_errorbar(aes(ymin=lowerall, ymax=upperall),width=0.5)+
49   geom_hline(yintercept =1, linetype=2)+
50
51   xlab('')+ ylab(" ")
52   ggtitle("BARI vs TNFi")+
53
54   scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
55 +
56
57
58
59
60

```

```

1
2
3   facet_wrap(~textall,nrow=16, strip.position= "right", scales =
4 "free_y") +
5
6   theme_pubclean()+
7   theme(strip.text.y = element_blank(),
8         strip.background = element_blank(),
9         axis.line.x = element_line(size = 0.5),
10        axis.text = element_text(face = "bold", colour = "black"),
11        legend.position="bottom", plot.margin =
12 unit(c(1,3,2,1),"lines"))+
13
14   coord_flip()
15
16

```

```

17 HR_plot_1
18

```

```

19 # adding some manual annotation
20 grid.text("Improves drug maintenance", x = unit(0.3, "npc"), y =
21 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
22 grid.text("Reduces drug maintenance", x = unit(0.87, "npc"), y =
23 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
24

```

Non-adjusted Kaplan-Meier curves

based on mini-tutorial found on datacamp.com/community/tutorials/survival-analysis-R

BARI vs TNFi

```

31 surv_object1 <- Surv(time = BARI1$time_on_drug, event =
32 BARI1$stop_DMARD) # indicate time on drug and stop variable
33 summary(coxph(surv_object1 ~ cohort, data=BARI1))
34 fit1 <- survfit(surv_object1 ~ cohort, data = BARI1) # this function
35 creates the data for Kaplan Meyer
36 fit1
37
38 survplot_1 <- ggsurvplot(fit1, data = BARI1, # plot
39   pval = T,
40   pval.method = TRUE,
41   legend.title = "Groups :",
42   legend.labs = c("Baricitinib", "TNFi"),
43   xlab = "Time (days)",
44   xlim = c(0, 700),
45   censor = FALSE,
46   title = "Non-adjusted model of drug discontinuation by type
47 of treatment",
48   surv.median.line = "v",
49   linetype = 1,
50   size = 1.5,
51   ggtheme = theme_minimal(),
52   #palette = c("grey78", "grey10"),
53   palette = c("red2", "green3"), # specify colors
54   risk.table = T)
55
56 survplot_1
57
58
59
60

```

```
1
2
3 summary(fit1, times = 365)
4 summary(fit1, times = 730)
5
6 Saving the plot curv object for Lilly
7
8 plot_BARI_vs_TNFi_data <- survplot_1$data.survplot
9 write.xlsx(plot_BARI_vs_TNFi_data, file =
10 ".\3_clean_output\Lilly_curves_excel\plot_BARI_vs_TNFi_data_non_adjust
11 ed.xlsx", row.names = F)
12
```

Home-made attempt to obtain adjusted curves based on imputed data

```
13 dummy_cox_impute1 <- mice::complete(imputed_data1, "long", include =
14 T)
15 dummy_cox_impute1 <- dummy_cox_impute1[dummy_cox_impute1$.imp != 0,]
16
17 BARI_fit1 <- survfit(coxph(Surv(time = time_on_drug, event =
18 stop_DMARD) ~ cohort+
19
20 I(age_base/10)+
21 bmi_base+
22 TC_with_csDMARD+
23 PREDNISON_STEROID+
24 CDAI0+
25 I(disease_duration_base_years/10)+
26 C(smoker_base, base=3)+
27 line_of_therapy+
28 gender+
29 seropositivity_base+
30 cluster(patient_id)+
31 strata(cohort), dummy_cox_impute1), data =
32 dummy_cox_impute1)
33
34
35
36
37 survplot_1_adj <- ggsurvplot(BARI_fit1, data = dummy_cox_impute1,
38 variable = "cohort",
39 xlab = "Time (days)",
40 title = "Multivariable Cox model of drug discontinuation by
41 type of treatment - BARI vs TNFi",
42 legend.title = "Groups :",
43 legend.labs = c("Baricitinib", "TNFi"),
44 censor = FALSE,
45 xlim = c(0, 700),
46 surv.median.line = "v",
47 linetype = 1,
48 size = 1.5,
49 ggtheme = theme_minimal(),
50 # palette = c("grey78", "grey10")
51 palette = c("red2", "green3") # to change colors
52 )
53
54
55 # adding some legends
56
57
58
59
60
```

```
1
2
3 survplot_1_adj <- survplot_1_adj +
4   labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
5 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
6 smoking status, line of therapy, gender, seropositivity")
7
```

```
8
9 survplot_1_adj
10 # summary(BARI_fit1) # to see detailed surv probabilities at given
11 timepoints
12 summary(BARI_fit1, times = 365)
13 summary(BARI_fit1, times = 730)
14
```

Saving the plot curv object for Lilly

```
15
16
17 plot_BARI_vs_TNFi_data_adj <- survplot_1_adj$data.survplot
18 write.xlsx(plot_BARI_vs_TNFi_data_adj, file =
19 ". /3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_adj.xlsx",
20 row.names = F)
21
```

Sensitivity analysis with package RiskRegression (AIPTW)

```
22
23 # Rappel: imputed_data1 = BARI vs TNFi
24 #   imputed_data2 = BARI vs OMA
25 library(riskRegression)
26
27 # I select only one imputed dataset. Would be even better to find a
28 way to pool/average the results from the 50 imputed datasets, but it
29 does not seem doable by default
30 test.data <- complete(imputed_data1, 1)
31
```

```
32 # First, we specify the treatment model (propensity score model)
33 # Logistic regression where the treatment group is the dependent
34 variable.
35
```

```
36
37 m.treatment <- glm(cohort~I(age_base/10)+
38                   bmi_base+
39                   TC_with_csDMARD+
40                   PREDNISON_STEROID+
41                   I(CDAI0/10)+
42                   I(disease_duration_base_years/10)+
43                   C(smoker_base, base=3)+
44                   line_of_therapy+
45                   gender+
46                   seropositivity_base,
47                   data = test.data, family =
48 "binomial" )
49
```

```
50
51 # Then we specify both the "event model" and the "censoring model".
52 Both are cox model
53
```

```
54 m.event <- coxph(Surv(time_on_drug, stop_DMARD) ~ cohort+
55                 I(age_base/10)+
56
57
58
59
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

bmi_base+
TC_with_csDMARD+
PREDNISON_STEROID+
I(CDAI0/10)+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base,
data = test.data, x = TRUE, y =
TRUE)

m.censor <- coxph(Surv(time_on_drug, stop_DMARD==0) ~ cohort +
I(age_base/10)+
bmi_base+
TC_with_csDMARD+
PREDNISON_STEROID+
I(CDAI0/10)+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base
, x =TRUE, y = TRUE,
data = test.data)

# And we measure the average treatment effect using function "ate",
specifying the time at which we want to compute the ATE

out <- ate(event = m.event ,
treatment = m.treatment,
censor = m.censor,
data = test.data,
cause = 1,
estimator = "AIPTW",
times = seq(from = 0, to = 500, by = 5))

dt.out <- as.data.table(out)

Diagnostics asked by Lily statistician

library(cobalt)

# First, the distribution of propensity scores
test.data$pscores <- m.treatment$fitted.values
test.data %>% setDT()

pscore_plot <- ggplot(test.data, aes(x = pscores, color = cohort, fill
= cohort)) +

```

```
1
2
3   geom_density(alpha = .47) +
4   xlab("Estimated Probability of being assigned BARI") +
5   ylab("Density") +
6   theme_minimal()+
7     theme(axis.ticks.y = element_blank(),
8           panel.grid.minor = element_blank(),
9           legend.title = element_blank(),
10          text = element_text(size = 16),
11          axis.title.x = element_text(hjust = 0.2, size = 16))
12
13 pscore_plot # overlap
14
15
16 ## Computing the weights
17 test.data$weights <- ifelse(test.data$cohort == "TNFi",
18 1/test.data$pscores, 1/(1-test.data$pscores))
19
20 # Selecting only our covariates of interest (the ones in the ps model)
21 COVS <- subset(test.data, select = c(cohort,age_base,
22                                   bmi_base,
23                                   TC_with_csDMARD,
24                                   PREDNISON_STEROID,
25                                   CDAI0,
26                                   disease_duration_base_years,
27                                   smoker_base,
28                                   line_of_therapy,
29                                   gender,
30                                   seropositivity_base))
31
32
33 # To get the SMD & variance ratios before/after weighting
34 # bal.tab(COVS, treat = test.data$cohort, thresholds = 0.1)
35 # bal.tab(COVS, treat = test.data$cohort, weights = test.data$weights,
36 thresholds = 0.1)
37 # bal.tab(COVS, treat = test.data$cohort, v.threshold = 2)
38 # bal.tab(COVS, treat = test.data$cohort, weights = test.data$weights,
39 v.threshold = 2)
40
41
42
43 # But plotting is clearer:
44 love.plot(COVS, treat = test.data$cohort, weights = test.data$weights,
45 stats = c("mean.diffs"), thresholds = c(m = .1), var.order =
46 "adjusted")
47
48
49 # We can also plot variance ratios for continuous variables
50 love.plot(COVS, treat = test.data$cohort, weights = test.data$weights,
51 stats = c("variance.ratios"))
52
53 # propensity scores enhance the balance overall, except for the CDAI0.
54 However, this is the reason we use the AIPTW. The remaining imbalance
55 is accounted for by the outcome model (outcome model is the cox
56
57
58
59
60
```


regression), and the misspecification of the outcome model is mitigated by the balancing done by propensity score.

First plot to get the difference in average treatment effect in percentage

```
plot.ate.diff <- ggplot(dt.out[type == "meanRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.4)+
  geom_line(aes(y = estimate, color = level), size = 1)+
  #geom_vline(xintercept = 90)+

  scale_colour_manual(values = c("lightblue","darkseagreen3"))+
  scale_fill_manual(values = c("lightblue","darkseagreen3"))+
  theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
legend.position="top" )+
  scale_x_continuous(breaks=seq(0,500,50)) + scale_y_continuous(labels
= scales::percent)+

  xlab("Days since initiation of treatment")+
  ylab("Absolute Risk of treatment discontinuation (%)")+
  labs(colour="Groups:", fill = "Groups:")+
  labs(group = "Groups:")+
  theme_bw(base_size = 14)+
  theme(axis.title.x = element_text(margin = margin(t = .3,unit =
"cm")),
axis.title.y = element_text(margin = margin(r = .3,unit =
"cm")))
```

plot.ate.diff

Second plot to get the ratio in average treatment effect

```
plot.ate.ratio <- ggplot(dt.out[type == "ratioRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.4)+
  geom_line(aes(y = estimate, color = level), size = 2)+

  theme_minimal()+
  theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
  scale_x_continuous(breaks=seq(0,500,50))+
  scale_y_continuous(limits = c(0.9,4.5))+

  xlab("Days since initiation of treatment")+
  ylab("Ratio in Average Treatment Effect")+
  labs(colour="treatment", fill = "treatment")
```

plot.ate.ratio

We can also consider the AIPTW estimate at a specific time point. For example at 365-day.

```

1
2
3   r.one <- dt.out[type == "diffRisk" & time == 365, .
4     (estimate,lower,upper,p.value)]
5   r.two <- dt.out[type == "ratioRisk" & time == 365, .
6     (estimate,lower,upper,p.value)]
7
8   ploufrows <- c("Difference in average treatment effect","Ratio in
9     average treatment effect")
10  ploufcols <- c("Estimate","95%CI","p")
11  table <- matrix(data = NA, nrow = length(ploufrows), ncol =
12    length(ploufcols))
13  rownames(table) <- ploufrows
14  colnames(table) <- ploufcols
15
16
17  library(formattable)
18  table[1,"Estimate"] <- paste0(formattable(r.one$estimate*100),"%")
19  table[1,"95%CI"] <-
20    paste0(formattable(r.one$lower),"-",formattable(r.one$upper))
21  table[1,"p"] <- writepvalue(r.one$p.value)
22  table[2,"Estimate"] <- paste0(r.two$estimate)
23  table[2,"95%CI"] <- paste0(r.two$lower,"-",r.two$upper)
24  table[2,"p"] <- writepvalue(r.two$p.value)
25
26
27  table
28
29  # Interpretation: If every patient had received BARI, the 365-day risk
30  of treatment discontinuation would have been 19.34% (points) lower
31  compared to when every patient had received TNFi.
32
33  BARI vs OMA ---
34
35  COX model
36  BARI2.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
37    cohort+
38      I(age_base/10)+
39      bmi_base+
40      TC_with_csDMARD+
41      PREDNISON_STEROID+
42      I(CDAI0/10)+
43      I(disease_duration_base_years/10)+
44      C(smoker_base, base=3)+
45      line_of_therapy+
46      gender+
47      seropositivity_base+
48      cluster(patient_id),
49      fitter = coxph, xtrans = imputed_data2,
50
51  data = BARI2)
52
53
54  summary(BARI2.adj.mi)
55
56
57
58
59
60

```

Creation of HR table and p-values (denis)

```

1
2
3
4 ploufrows <- names(BARI2.adj.mi$coefficients)
5 ploufcols <- c("HR","95%CI","p")
6 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
7 length(ploufcols))
8 rownames(coxtable) <- ploufrows
9 colnames(coxtable) <- ploufcols
10 plouf <- summary(BARI2.adj.mi)
11
12
13 for(row in ploufrows)
14 {
15   coxtable[row,"HR"] <-
16   formattable(plouf$coefficients[row,"exp(coef)"])
17   coxtable[row,"95%CI"] <-
18   paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",formattable(pl
19 ouf$conf.int[row,"upper .95"]))
20   coxtable[row,"p"] <- writepvalue(plouf$coefficients[row,"Pr(>|
21 z|)"])
22 }
23
24
25 write.xlsx(coxtable, file="./3_clean_output/BARI vs OMA HR.xlsx") #
26 saving excel file
27

```

Forest plot

```

28
29 meanall <- summary(BARI2.adj.mi)$coefficients[1:14,"exp(coef)"]
30 lowerall <- summary(BARI2.adj.mi)$conf.int[1:14,"lower .95"]
31 upperall <- summary(BARI2.adj.mi)$conf.int[1:14,"upper .95"]
32 textall <- c("OMA (vs BARI)", "Age (decades)", "BMI", "Concomitant
33 csDMARD", "Concomitant glucocorticoid", "CDAI score (10 pts)",
34 "Disease duration (decades)", "Current smoker (vs non-smoker)",
35 "Former smoker (vs non-smoker)", "2nd line therapy (vs
36 1st)", "3rd line therapy (vs 1st)", "4th or later line (vs 1st)",
37 "Female gender", "Seropositivity (RF or ACPA)")
38
39
40 dfall <- data.frame(textall, meanall, lowerall, upperall)
41 dfall$textall <- factor(dfall$textall,
42 levels = textall)
43
44 HR_plot_2 <- ggplot(data=dfall, aes(x=textall, y= meanall, ymin =
45 lowerall, ymax = upperall))+
46
47   geom_pointrange(size=0.5)+
48   geom_errorbar(aes(ymin=lowerall, ymax=upperall),width=0.5)+
49   geom_hline(yintercept =1, linetype=2)+
50
51   xlab('')+ ylab(" ")
52   ggtitle("BARI vs OMA")+
53
54   scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
55 +
56
57
58
59
60

```

```

1
2
3   facet_wrap(~textall,nrow=16, strip.position= "right", scales =
4 "free_y") +
5
6   theme_pubclean()+
7   theme(strip.text.y = element_blank(),
8         strip.background = element_blank(),
9         axis.line.x = element_line(size = 0.5),
10        axis.text = element_text(face = "bold", colour = "black"),
11        legend.position="bottom", plot.margin =
12 unit(c(1,3,2,1),"lines"))+
13
14
15     coord_flip()
16
17 HR_plot_2
18
19 # adding some manual annotation
20 grid.text("Improves drug maintenance", x = unit(0.3, "npc"), y =
21 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
22 grid.text("Reduces drug maintenance", x = unit(0.87, "npc"), y =
23 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
24

```

Non-adjusted Kaplan-Meier curves

based on mini-tutorial found on datacamp.com/community/tutorials/survival-analysis-R)

BARI vs OMA

```

31 surv_object2 <- Surv(time = BARI2$time_on_drug, event =
32 BARI2$stop_DMARD)
33 fit2 <- survfit(surv_object2 ~ cohort, data = BARI2) # this function
34 creates the data for Kaplan Meyer
35 survplot_2 <- ggsurvplot(fit2, data = BARI2, # plot
36   pval = T,
37   pval.method = TRUE,
38   legend.title = "Groups :",
39   legend.labs = c("Baricitinib", "OMA"),
40   xlab = "Time (days)",
41   xlim = c(0, 700),
42   censor = FALSE,
43   title = "Non-adjusted model of drug discontinuation by type
44 of treatment",
45   surv.median.line = "v",
46   linetype = 1,
47   size = 1.5,
48   ggtheme = theme_minimal(),
49   #palette = c("grey78", "grey50"),
50   palette = c("red2", "blue3"), # to put colors
51   risk.table = T)
52
53 survplot_2
54 summary(fit2, times = 365)
55 summary(fit2, times = 730)
56
57
58
59
60

```

Saving the plot curv object for Lilly

```
plot_BARI_vs_OMA_data <- survplot_2$data.survplot
write.xlsx(plot_BARI_vs_OMA_data, file =
"/3_clean_output/Lilly_curves_excel/plot_BARI_vs_OMA_data_non_adjuste
d.xlsx", row.names = F)
```

Home-made attempt to obtain adjusted cuves based on imputed data :

```
dummy_cox_impute2 <- mice::complete(imputed_data2, "long", include =
T)
```

```
dummy_cox_impute2 <- dummy_cox_impute2[dummy_cox_impute2$.imp != 0,]
```

```
BARI_fit2 <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+
```

```
    I(age_base/10)+
```

```
    bmi_base+
```

```
    TC_with_csDMARD+
```

```
    PREDNISON_STEROID+
```

```
    CDAI0+
```

```
    I(disease_duration_base_years/10)+
```

```
    C(smoker_base, base=3)+
```

```
    line_of_therapy+
```

```
    gender+
```

```
    seropositivity_base+
```

```
    cluster(patient_id)+
```

```
    strata(cohort),dummy_cox_impute2), data =
```

```
dummy_cox_impute2)
```

```
survplot_2_adj <- ggsurvplot(BARI_fit2, data = dummy_cox_impute2,
variable = "cohort",
```

```
    xlab = "Time (days)",
```

```
    title = "Multivariable Cox model of drug discontinuation by
```

```
type of treatment - BARI vs OMA",
```

```
    legend.title = "Groups :",
```

```
    legend.labs = c("Baricitinib", "OMA bDMARDs"),
```

```
    censor = FALSE,
```

```
    xlim = c(0, 700),
```

```
    surv.median.line = "v",
```

```
    linetype = 1,
```

```
    size = 1.5,
```

```
    ggtheme = theme_minimal(),
```

```
    # palette = c("grey78", "grey50")
```

```
    palette = c("red2", "blue3") # to change colors
```

```
)
```

```
# adding some legends
```

```
survplot_2_adj <- survplot_2_adj +
```

```
  labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
concomitant glucocorticoid, baseline CDAI, disease duration (decades),
smoking status, line of therapy, gender, seropositivity")
```

```

1
2
3
4  survplot_2_adj
5  summary(BARI_fit2, times = 365) # to see detailed surv probabilities
6  at given timepoints
7  summary(BARI_fit2, times = 730)
8

```

Saving the plot curv object for Lilly

```

9
10
11  plot_BARI_vs_OMA_data_adj <- survplot_2_adj$data.survplot
12  write.xlsx(plot_BARI_vs_OMA_data_adj, file =
13  ".\3_clean_output\Lilly_curves_excel\plot_BARI_vs_OMA_data_adj.xlsx",
14  row.names = F)
15

```

Sensitivity analysis with package RiskRegression (AIPTW)

I select only one imputed dataset. Would be good to find a way to pool the results from the 50 datasets imputed.

```

16
17
18  test.data2 <- complete(imputed_data2,1)
19
20

```

First, we specify the treatment model (propensity score model)
Logistic regression where the treatment group is the dependent variable.

```

21
22
23  m.treatment2 <- glm(cohort~I(age_base/10)+
24  bmi_base+
25  TC_with_csDMARD+
26  PREDNISON_STEROID+
27  I(CDAI0/10)+
28  I(disease_duration_base_years/10)+
29  C(smoker_base, base=3)+
30  line_of_therapy+
31  gender+
32  seropositivity_base,
33  data = test.data2, family =
34  "binomial" )
35

```

Then we specify both the "event model" and the "censoring model".
Both are cox model

```

36
37
38
39
40
41  m.event2 <- coxph(Surv(time_on_drug, stop_DMARD) ~ cohort+
42  I(age_base/10)+
43  bmi_base+
44  TC_with_csDMARD+
45  PREDNISON_STEROID+
46  I(CDAI0/10)+
47  I(disease_duration_base_years/10)+
48  C(smoker_base, base=3)+
49  line_of_therapy+
50  gender+
51  seropositivity_base,
52  data = test.data2, x = TRUE, y =
53
54
55
56
57
58
59
60

```

```
1
2
3 TRUE)
4
5 m.censor2 <- coxph(Surv(time_on_drug,stop_DMARD==0) ~ cohort +
6 I(age_base/10)+
7 bmi_base+
8 TC_with_csDMARD+
9 PREDNISOLONE_STEROID+
10 I(CDAI0/10)+
11 I(disease_duration_base_years/10)+
12 C(smoker_base, base=3)+
13 line_of_therapy+
14 gender+
15 seropositivity_base
16 , x =TRUE, y = TRUE,
17 data = test.data2)
18
19
20 # And we measure the average treatment effect using function "ate",
21 specifying the times at which we want to compute the ATE
22
23
24 out2 <- ate(event = m.event2 ,
25 treatment = m.treatment2,
26 censor = m.censor2,
27 data = test.data2,
28 cause = 1,
29 estimator = "AIPTW",
30 times = seq(from = 0, to = 500, by = 5))
31
32
33 dt.out2 <- as.data.table(out2)
34
35 Diagnostics asked by Lily statistician
36
37 library(cobalt)
38
39 # First, the distribution of propensity scores
40 test.data2$pscores <- m.treatment2$fitted.values
41 test.data2 %>% setDT()
42
43 pscore_plot2 <- ggplot(test.data2,aes(x = pscores, color = cohort,
44 fill = cohort)) +
45 geom_density(alpha = .47) +
46 xlab("Estimated Probability of being assigned BARI") +
47 ylab("Density") +
48 theme_minimal()+
49 theme(axis.ticks.y=element_blank(),
50 panel.grid.minor=element_blank(),
51 legend.title=element_blank(),
52 text = element_text(size = 16),
53 axis.title.x =element_text(hjust = 0.2, size = 16))
54 pscore_plot2
55 # Good overlap
56
57
58
59
60
```

```
1
2
3
4 ## Computing the weights
5 test.data2$weights <- ifelse(test.data2$cohort == "OMA",
6 1/test.data2$pscores, 1/(1-test.data2$pscores))
7
8
9 # Selecting only our covariates of interest (the ones in the ps model)
10 COVS_2 <- subset(test.data2, select = c(cohort,age_base,
11 bmi_base,
12 TC_with_csDMARD,
13 PREDNISON_STEROID,
14 CDAI0,
15 disease_duration_base_years,
16 smoker_base,
17 line_of_therapy,
18 gender,
19 seropositivity_base))
20
21
22 # To get the SMD & variance ratios before/after weighting
23 # bal.tab(COVS_2, treat = test.data2$cohort, thresholds = 0.1)
24 # bal.tab(COVS_2, treat = test.data2$cohort, weights =
25 test.data2$weights, thresholds = 0.1)
26 # bal.tab(COVS_2, treat = test.data2$cohort, v.threshold = 2)
27 # bal.tab(COVS_2, treat = test.data2$cohort, weights =
28 test.data2$weights, v.threshold = 2)
29 #
30
31 #But plotting it is better:
32 love.plot(COVS_2, treat = test.data2$cohort, weights =
33 test.data2$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
34 var.order = "adjusted")
35
36
37 # We can also plot variance ratios for continuous variables
38 love.plot(COVS_2, treat = test.data2$cohort, weights =
39 test.data2$weights,stats = c("variance.ratios"))
40
41 # propensity scores enhance the balance overall, except for the CDAI0.
42 However, this is the reason we use the AIPTW. The remaining imbalance
43 is accounted for by the outcome model (outcome model is the cox
44 regression), and the misspecification of the outcome model is
45 mitigated by the balancing done by propensity score.
46
47 First plot to get the difference in average treatment effect in percentage
48
49 plot.ate.diff2 <- ggplot(dt.out2[type == "meanRisk"], aes(x = time,
50 group = level))+
51   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
52 0.3)+
53   geom_line(aes(y = estimate, color = level), size = 1)+
54
55   theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
56
57
58
59
```



```

1
2
3 legend.position="top" )+
4   scale_x_continuous(breaks=seq(0,500,50)) + scale_y_continuous(labels
5 = scales::percent)+
6
7   xlab("Days since initiation of treatment")+
8   ylab("Absolute Risk of treatment discontinuation (%)")+
9   labs(colour="Groups:", fill = "Groups:", title = "Absolute risk of
10 treatment discontinuation by type of treatment - BARI vs TNFi")+
11   labs(group = "Groups:")
12
13

```

```
14 plot.ate.diff2
```

15 Second plot to get the ratio in average treatment effect

```

16
17 plot.ate.ratio2 <- ggplot(dt.out2[type == "ratioRisk"], aes(x = time,
18 group = level))+
19   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
20 0.3)+
21   geom_line(aes(y = estimate, color = level), size = 1)+
22
23
24   theme_minimal()+
25   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
26   scale_x_continuous(breaks=seq(100,400,50))+
27   scale_y_continuous(limits = c(0.8,3))+
28
29   xlab("Days since initiation of treatment")+
30   ylab("Ratio in Average Treatment Effect")+
31   labs(colour="treatment", fill = "treatment")
32
33

```

```
34 plot.ate.ratio2
```

35 We can also consider the AIPTW estimate at a specific time point. For example at 365-day.

```

36
37 r.one <- dt.out2[type == "diffRisk" & time == 365, .
38 (estimate,lower,upper,p.value)]
39 r.two <- dt.out2[type == "ratioRisk" & time == 365, .
40 (estimate,lower,upper,p.value)]
41
42
43 ploufrows <- c("Difference in average treatment effect","Ratio in
44 average treatment effect")
45 ploufcols <- c("Estimate","95%CI","p")
46 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
47 length(ploufcols))
48 rownames(coxtable) <- ploufrows
49 colnames(coxtable) <- ploufcols
50
51 library(formattable)
52 coxtable[1,"Estimate"] <- paste0(formattable(r.one$estimate*100),"%")
53 coxtable[1,"95%CI"] <-
54 paste0(formattable(r.one$lower),"-",formattable(r.one$upper))
55 coxtable[1,"p"] <- writepvalue(r.one$p.value)
56
57
58
59
60

```

```

1
2
3 coxtable[2,"Estimate"] <- paste0(r.two$estimate)
4 coxtable[2,"95%CI"] <- paste0(r.two$lower,"-",r.two$upper)
5 coxtable[2,"p"] <- writepvalue(r.two$p.value)
6
7
8 coxtable
9
10 # Interpretation: If every patient had received BARI, the 365-day risk
11 of treatment discontinuation would have been xx% (points) lower
12 compared to when every patient had received TNFi.
13
14
15
16
17
18

```

[3] 1st LINE vs 1st LINE analysis

Common Table 1

Table 1 with NA, to have exact counts and proportions

```

25 BARI_first <- BARI_DATA
26 BARI_first <- BARI_first[line_of_therapy == "1st"] # selection of TC
27 first line
28
29 myVars2 <- c("gender", "age_base", "disease_duration_base_years",
30 "CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
31 "seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
32 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
33 "PREDNISON_STEROID_dose", "dose", "initiation_year",
34 "time_on_drug", "HAQ_score_base")
35
36
37 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
38 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
39 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
40 "seropositivity_base")
41
42
43 nonnormalVars <- c()
44
45 tab1 <- CreateTableOne(vars = myVars2, data = BARI_first, factorVars =
46 catVars2, strata = "cohort", test = F, includeNA = T)
47 tablexp <- print(tab1, nonnormal= nonnormalVars, catDigits = 1,
48 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
49
50 saving
51
52 write.xlsx(tablexp, file = "./3_clean_output/BARI 3 groups first line
53 table1 NA.xlsx")
54

```

Table 1 without NA to have adequate p values to interpret

```
1
2
3 BARI_first <- BARI_DATA
4 BARI_first <- BARI_first[line_of_therapy == "1st"] # selection TC
5 first line
6
7
8 myVars2 <- c("gender", "age_base", "disease_duration_base_years",
9 "CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
10 "seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
11 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
12 "PREDNISON_STEROID_dose", "dose", "initiation_year",
13 "time_on_drug", "HAQ_score_base")
14
15 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
16 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
17 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
18 "seropositivity_base")
19
20 nonnormalVars <- c()
21
22
23 tab1 <- CreateTableOne(vars = myVars2, data = BARI_first, factorVars =
24 catVars2, strata = "cohort", test = T, includeNA = F)
25 tablexp <- print(tab1, nonnormal= nonnormalVars, catDigits = 1,
26 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
27
28 Saving
29
30 write.xlsx(tablexp, file = "./3_clean_output/BARI 3 groups first line
31 table1.xlsx")
32
33 summary(BARI_first[cohort=="BARI", c("TC_id", "patient_id",
34 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
35 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
36 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
37 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
38 "cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
39 to see NA values for all variables
40
41 summary(BARI_first[cohort=="TNFi", c("TC_id", "patient_id",
42 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
43 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
44 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
45 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
46 "cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
47 to see NA values for all variables
48
49
50 summary(BARI_first[cohort=="OMA", c("TC_id", "patient_id",
51 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
52 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
53 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
54 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
55
56
57
58
59
60
```

```
1
2
3 "cohort", "adverse_event_reported", "seropositivity_base", "dose"]]) #
4 to see NA values for all variables
5
```

6 Non-adjusted Survival curves

8 BARI vs TNFi

```
9
10 BARI_first1 <- copy(BARI_first[cohort %in% c("BARI", "TNFi")])
11
12 surv_object3 <- Surv(time = BARI_first1$time_on_drug, event =
13 BARI_first1$stop_DMARD) # indicate stop variable and time_on_drug
14 summary(coxph(surv_object3 ~ cohort, data = BARI_first1))
15 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first1) # function
16 which creates Kaplan-meier data
17 survplot_first1 <- ggsurvplot(fit3, data = BARI_first1, # plot
18 pval = T,
19 pval.method = TRUE,
20 legend.title = "Groups :",
21 legend.labs = c("Baricitinib", "TFNi"),
22 xlab = "Time (days)",
23 xlim = c(0, 700),
24 censor = FALSE,
25 title = "Non-adjusted model of drug discontinuation by type
26 of treatment",
27 surv.median.line = "v",
28 linetype = 1,
29 size = 1.5,
30 ggtheme = theme_minimal(),
31 # palette = c("grey78", "grey50", "grey10"),
32 palette = c("red2", "green3"), # to get colors
33 risk.table = T
34 )
35
36
37
```

```
38 survplot_first1
39 table(BARI_first1$cohort)
40 summary(fit3)
41
```

```
42 rm(surv_object3, fit3)
43
```

44 BARI vs OMA

```
45
46 BARI_first2 <- BARI_first[line_of_therapy == "1st" & cohort %in%
47 c("BARI", "OMA")] # selection des TC TNFi
48
```

```
49
50 surv_object3 <- Surv(time = BARI_first2$time_on_drug, event =
51 BARI_first2$stop_DMARD) # indicate stop variable and time_on_drug
52 summary(coxph(surv_object3 ~ cohort, data = BARI_first2))
53 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first2) # function
54 which creates Kaplan-meier data
55 survplot_first2 <- ggsurvplot(fit3, data = BARI_first2, # plot
56 pval = T,
57
58
59
```

```

1
2
3         pval.method = TRUE,
4         legend.title = "Groups :",
5         legend.labs = c("Baricitinib", "OMA"),
6         xlab = "Time (days)",
7         xlim = c(0, 700),
8         censor = FALSE,
9         title = "Non-adjusted model of drug discontinuation by type
10 of treatment",
11         surv.median.line = "v",
12         linetype = 1,
13         size = 1.5,
14         ggtheme = theme_minimal(),
15         # palette = c("grey78", "grey50", "grey10"),
16         palette = c("red2", "blue3"), # to get colors
17         risk.table = T
18     )
19
20
21 survplot_first2
22 table(BARI_first2$cohort)
23 summary(fit3)

```

```

24
25
26 rm(surv_object3, fit3)

```

Adjusted survival analyses

BARI vs TNFi

Verification (quick)

```

31
32 # Test of proportionality of hazards on raw data
33 test_first_ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
34 as.factor(cohort)+
35
36
37         cluster(patient_id),
38         data= BARI_first1)
39 cox.zph(test_first_ph)
40

```

Adjusted Cox-model

```

41
42 imputed_data1_first <- complete(imputed_data1,"long",include=T) # to
43 put in the long format
44 imputed_data1_first <- filter(imputed_data1_first, line_of_therapy ==
45 "1st") # only keep 1st line imputed TC
46 imputed_data1_first <- as.mids(imputed_data1_first) # put back in
47 previous format, to use fit.mult.impute
48
49
50 BARI_first1.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
51 cohort+
52
53         I(age_base/10)+
54         bmi_base+
55         concomitant_csDMARD+
56         PREDNISON_STEROID+
57
58
59

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

I(CDAI0/10)+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base+
cluster(patient_id),
fitter = coxph, xtrans =

```

imputed_data1_first, data = BARI_first1)
summary(BARI_first1.adj.mi)

```

Creation of HR table with p-values

```

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

ploufrows <- names(BARI_first1.adj.mi$coefficients)
ploufcols <- c("HR", "95%CI", "p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols
plouf <- summary(BARI_first1.adj.mi)

for(row in ploufrows)
{
  coxtable[row, "HR"] <-
formattable(plouf$coefficients[row, "exp(coef)"])
  coxtable[row, "95%CI"] <-
paste0(formattable(plouf$conf.int[row, "lower .95"]), "-", formattable(pl
ouf$conf.int[row, "upper .95"]))
  coxtable[row, "p"] <- writepvalue(plouf$coefficients[row, "Pr(>|
z|)"])
}

write.xlsx(coxtable, file="./3_clean_output/BARI vs TNFi HR 1st
lines.xlsx") # save in excel format

```

Adjusted curves with imputed data

```

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

dummy_cox_impute_first1 <- mice::complete(imputed_data1_first, "long",
include = T)
dummy_cox_impute_first1 <-
dummy_cox_impute_first1[dummy_cox_impute_first1$.imp != 0,]

```

```

BARI_first1_fit <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+

```

I(age_base/10)+
bmi_base+
concomitant_csDMARD+
PREDNISON_STEROID+
CDAI0+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+

```

1
2
3           gender+
4           seropositivity_base+
5           cluster(patient_id)+
6           strata(cohort), dummy_cox_impute_first1),
7 data = dummy_cox_impute_first1)
8
9
10
11 survplot_first1_adj <- ggsurvplot(BARI_first1_fit, data =
12 dummy_cox_impute_first1, variable = "cohort",
13   xlab = "Time (days)",
14   title = "Multivariable Cox model of drug discontinuation by
15 type of treatment - 1st line vs 1st line",
16   legend.title = "Groups :",
17   legend.labs = c("Baricitinib", "TNFi"),
18   censor = FALSE,
19   xlim = c(0, 700),
20   surv.median.line = "v",
21   linetype = 1,
22   size = 1.5,
23   ggtheme = theme_minimal(),
24   #palette = c("grey78", "grey10")
25   palette = c("red2", "green3"), # to get colors
26 )
27
28
29 survplot_first1_adj <- survplot_first1_adj +
30   labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
31 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
32 smoking status, line of therapy, gender, seropositivity")
33
34 survplot_first1_adj
35 table(BARI_first1$cohort)
36 rm(dummy_cox_impute_first1, BARI_first1_fit)

```

BARI vs OMA

Verification (quick)

```

42 # Test of proportionality of hazards on raw data
43 test_first_ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
44 as.factor(cohort)+
45   cluster(patient_id),
46   data= BARI_first2)
47
48 cox.zph(test_first_ph)

```

Adjusted Cox-model

```

51 imputed_data2_first <- complete(imputed_data2,"long",include=T) # to
52 put in the long format
53 imputed_data2_first <- filter(imputed_data2_first, line_of_therapy ==
54 "1st") # only keep 1st line imputed TC
55 imputed_data2_first <- as.mids(imputed_data2_first) # put back in
56
57
58
59

```

previous format, to use fit.mult.impute

```
BARI_first2.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
cohort+
```

```
      I(age_base/10)+
      bmi_base+
      concomitant_csDMARD+
      PREDNISON_STEROID+
      I(CDAI0/10)+
      I(disease_duration_base_years/10)+
      C(smoker_base, base=3)+
      line_of_therapy+
      gender+
      seropositivity_base+
      cluster(patient_id),
      fitter = coxph, xtrans =
```

```
imputed_data2_first, data = BARI_first2)
summary(BARI_first2.adj.mi)
```

Creation of HR table with p-values

```
ploufrows <- names(BARI_first2.adj.mi$coefficients)
ploufcols <- c("HR", "95%CI", "p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols
plouf <- summary(BARI_first2.adj.mi)

for(row in ploufrows)
{
  coxtable[row, "HR"] <-
formattable(plouf$coefficients[row, "exp(coef)"])
  coxtable[row, "95%CI"] <-
paste0(formattable(plouf$conf.int[row, "lower .95"]), "-", formattable(pl
ouf$conf.int[row, "upper .95"]))
  coxtable[row, "p"] <- writepvalue(plouf$coefficients[row, "Pr(>|
z|)"])
}
```

```
write.xlsx(coxtable, file="./3_clean_output/BARI vs OMA HR 1st
lines.xlsx") # save in excel format
```

Adjusted curves with imputed data

```
dummy_cox_impute_first2 <- mice::complete(imputed_data2_first, "long",
include = T)
dummy_cox_impute_first2 <-
dummy_cox_impute_first2[dummy_cox_impute_first2$.imp != 0,]

BARI_first2_fit <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+
```



```
1
2
3           I(age_base/10)+
4           bmi_base+
5           concomitant_csDMARD+
6           PREDNISON_STEROID+
7           CDAI0+
8           I(disease_duration_base_years/10)+
9           C(smoker_base, base=3)+
10          line_of_therapy+
11          gender+
12          seropositivity_base+
13          cluster(patient_id)+
14          strata(cohort), dummy_cox_impute_first2),
15 data = dummy_cox_impute_first2)
16
17
18
19 survplot_first2_adj <- ggsurvplot(BARI_first2_fit, data =
20 dummy_cox_impute_first2, variable = "cohort",
21   xlab = "Time (days)",
22   title = "Multivariable Cox model of drug discontinuation by
23 type of treatment - 1st line vs 1st line",
24   legend.title = "Groups :",
25   legend.labs = c("Baricitinib", "OMA bDMARDs"),
26   censor = FALSE,
27   xlim = c(0, 700),
28   surv.median.line = "v",
29   linetype = 1,
30   size = 1.5,
31   ggtheme = theme_minimal(),
32   #palette = c("grey78", "grey50")
33   palette = c("red2", "blue3"), # to get colors
34 )
35
36
37 survplot_first2_adj <- survplot_first2_adj +
38   labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
39 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
40 smoking status, line of therapy, gender, seropositivity")
41
42
43 survplot_first2_adj
44 table(BARI_first2$cohort)
45 rm(dummy_cox_impute_first2, BARI_first2_fit)
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

1. [4] LACK of EFFICACY and ADVERSE EVENTS

Analysis by stop_reasons in competing risk

(BARI vs TNFi)

Cumulative incidence function

```
BARI_comp <- copy(BARI_DATA)
```

```
#General
```

```
BARI_comp[stop_reasons == "ADVERSE_EVENT",status := 1]
BARI_comp[stop_reasons == "NOT_EFFECTIVE", status := 2]
BARI_comp[stop_reasons == "OTHER" | stop_reasons == "REMISSION",
status := 3]
BARI_comp[stop_reasons == "CONTINUE", status := 0]
BARI_comp$cohort <- as.factor(BARI_comp$cohort)
```

```
library(reshape)
```

```
BARI_comp_B <- BARI_comp[cohort %in% c("BARI")] #BARI only
ci_BARI <- Cuminc(time = "time_on_drug",status = "status", data =
BARI_comp_B)
ci_BARI <- ci_BARI[, -c(2,6,7,8,9)]
ci_long_BARI <- reshape2::melt(ci_BARI,id.vars = "time")
```

```
BARI_comp_T <- BARI_comp[cohort %in% c("TNFi")] #TNFi only
ci_TNFi <- Cuminc(time = "time_on_drug",status = "status", data =
BARI_comp_T)
ci_TNFi <- ci_TNFi[, -c(2,6,7,8,9)]
ci_long_TNFi <- reshape2::melt(ci_TNFi,id.vars = "time")
```

```
ci_long_BARI$cohort <- 0
ci_long_TNFi$cohort <- 1
ci_long <- rbind(ci_long_BARI,ci_long_TNFi)
ci_long$cohort <- as.factor(ci_long$cohort)
```

```
plot2 <- ggplot(data = ci_long, aes(x = time,
y = value,
linetype =
interaction(cohort,variable),
col =
interaction(cohort,variable))) +
  geom_line(size = 0.75) +
  scale_color_manual(name = "",
values
=c("#08306B", "#08306B", "#238B45", "#238B45", "#FD8D3C", "#FD8D3C"),
labels = c("Adverse Event (BARI)", "Adverse Event
(TNFi)", "Ineffectiveness (BARI)", "Ineffectiveness (TNFi)", "Other
```

```

1
2
3 (BARI)", "Other (TNFi)") +
4   scale_linetype_manual(name="",
5                         values = c(1,3,1,3,1,3),
6                         labels = c("Adverse Event (BARI)", "Adverse
7 Event (TNFi)", "Ineffectiveness (BARI)", "Ineffectiveness (TNFi)", "Other
8 (BARI)", "Other (TNFi)") +
9   scale_x_continuous(name = "Time", limits = c(1,365)) +
10  scale_y_continuous(name = "Cumulative incidence", limits =
11 c(0.0,0.3)) +
12  theme_bw() +
13  theme(strip.text.y = element_blank(),
14        strip.background = element_blank(),
15        axis.line.x = element_line(size = 0.5),
16        axis.text = element_text(face = "bold", colour = "black"),
17        legend.position="right", plot.margin =
18 unit(c(1,3,2,1), "lines")) +
19  #ggtitle("Cumulative incidence functions") +
20  theme(plot.title = element_text(hjust = 0.5))
21
22

```

```

23 plot2
24

```

Adjusting variables

```

25
26 # Covariates of interest for Cox
27

```

```

28
29 covs <-
30 c("cohort", "age_base", "bmi_base", "TC_with_csDMARD", "PREDNISON_STEROID"
31 , "CDAI0", "disease_duration_base_years", "smoker_base", "line_of_therapy"
32 , "gender", "seropositivity_base")
33

```

Cause-specific hazard model

```

34
35 # Rappel: imputed_data1 = BARI vs TNFi
36 #           imputed_data2 = BARI vs OMA
37

```

```

38 # Transition matrix definition
39

```

```

40 tmat <- trans.comprisk(2, names = c("event-free", "ae", "lae"))
41 tmat
42

```

```

43 imputed_data1_long <- complete(imputed_data1, action = "long") %>%
44 setDT()
45 imputed_data1_long[, stop_ae := fifelse(stop_reasons ==
46 "ADVERSE_EVENT", 1, 0)]
47 imputed_data1_long[, stop_lae := fifelse(stop_reasons ==
48 "NOT_EFFECTIVE", 1, 0)]
49 imputed_data1_long[, stop_other := fifelse(stop_reasons == "OTHER" |
50 stop_reasons == "REMISSION", 1, 0)]
51 #[, continue := fifelse(stop_reasons == "CONTINUE", 1, 0)]
52 imputed_data1_long[, continue := fifelse(stop_reasons == "OTHER" |
53 stop_reasons == "REMISSION" | stop_reasons == "CONTINUE", 1, 0)]
54

```

```

55
56 M <- imputed_data1$m
57
58
59

```

```
1
2
3
4 mice_fit <- lapply(1:M,function(m){
5
6   # subset
7   data_sub <- imputed_data1_long[.imp == m]
8
9   mst_hosp <- msprep(time =
10  c("time_on_drug","time_on_drug","time_on_drug"),
11      status = c("continue","stop_ae","stop_lae"),
12      data = as.data.frame(data_sub),
13      trans = tmat,
14      keep = covs)
15
16   # get covariates
17   tmp <- expand.covs(mst_hosp,covs, append = TRUE, longnames = T)
18   tmp_cov <- grep(paste0(covs,".",collapse = "|"),names(tmp),value = T)
19
20   # fit
21   coxph(as.formula(paste0("Surv(Tstart, Tstop, status) ~",
22      paste0(tmp_cov,collapse = " + "),
23      "+ strata(trans)")),
24      data = tmp,
25      method = "breslow")
26
27
28   }) %>%
29   as.mira()
30
31   est <- pool(mice_fit)
32
33   # Transition 1 = Adverse Event. Hazard Ratio vs VARI
34   # Transition 2 = Lack of Efficacy. Hazard Ratio vs VARI
35   # estimate = Hazard ratio
36
37
38   summary(est, conf.int = T, exponentiate = T)
39
40   # Conclusion
41   # => The hazard ratio of lack of efficacy (lae) for TNFi is 65% higher
42   than for BARI. Significant.
43   # => No difference between TNFi and BARI for Adverse Event (ae)
44
45   Clean table with confidence intervals & p-values
46
47   # Hazard ratios
48   ploufrows <- as.character(summary(est)$term)
49
50   ploufcols <- c("HR","95%CI","p")
51   coxtable_csh <- matrix(data = NA, nrow = length(ploufrows), ncol =
52   length(ploufcols))
53   rownames(coxtable_csh) <- ploufrows
54   colnames(coxtable_csh) <- ploufcols
55   plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
```

```

1
2
3
4   for(row in ploufrows)
5   {
6     coxtable_csh[row,"HR"] <- formattable(plouf[term %in% row,
7     estimate])
8     coxtable_csh[row,"95%CI"] <- paste0(formattable(plouf[term %in%
9     row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
10    coxtable_csh[row,"p"] <- writepvalue(plouf[term %in% row,
11    p.value])}
12
13
14   output <- coxtable_csh
15   row.names(output)[1:2] <- c("TNFi Adverse Event (vs BARI)", "TNFi Lack
16   of Eff (vs BARI)")
17
18   # Transition 1 = Adverse Event. Hazard Ratio vs VARI
19   # Transition 2 = Lack of Efficacy. Hazard Ratio vs VARI
20   output
21
22   Subdistribution hazard model (Fine-Gray)
23   # Status variable
24   imputed_data1_long[stop_reasons == "ADVERSE_EVENT",status := 1]
25   imputed_data1_long[stop_reasons == "NOT_EFFECTIVE", status := 2]
26   imputed_data1_long[stop_reasons == "OTHER" | stop_reasons ==
27   "REMISSION", status := 3]
28   imputed_data1_long[stop_reasons == "CONTINUE", status := 0]
29
30
31   ## ATTENTION levels() re-ecrit juste l'étiquette!! Change pas la
32   donnée !!! Donc ça re écrit les labels
33
34   imputed_data1_long$line_of_therapy <-
35   as.factor(imputed_data1_long$line_of_therapy)
36   imputed_data1_long$seropositivity_base <-
37   as.factor(imputed_data1_long$seropositivity_base)
38
39
40   levels(imputed_data1_long$cohort) <- c("0","1")
41   levels(imputed_data1_long$line_of_therapy) <- c("0","1","2","3")
42   levels(imputed_data1_long$gender) <- c("0","1")
43   levels(imputed_data1_long$smoker_base) <- c("2","1","0")
44   levels(imputed_data1_long$smoker_base)
45   levels(imputed_data1_long$seropositivity_base) <- c("0","1")
46
47   M <- imputed_data1$m
48
49
50   # First loop to get estimates for event = 1: ADVERSE EVENT
51
52   mice_fit <- lapply(1:M,function(m){
53     # subset
54     BARI_toto <- imputed_data1_long[.imp == m]
55     # subdistribution hazard model
56
57
58
59
60

```

```

1
2
3     shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
4 BARI_toto[,..covs],failcode = 1,cencode = 0)
5
6   }) %>%
7     as.mira()
8   est <- pool(mice_fit)
9   summary(est, conf.int = T, exponentiate = T)
10
11 # Second loop to get estimates for event = 2: LACK OF EFFICACY
12
13
14 mice_fit2 <- lapply(1:M,function(m){
15   # subset
16   BARI_toto <- imputed_data1_long[.imp == m]
17   #subdistribution hazard model
18   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
19 BARI_toto[,..covs],failcode = 2,cencode = 0)
20
21 }) %>%
22   as.mira()
23
24
25 est2 <- pool(mice_fit2)
26 summary(est2, conf.int = T, exponentiate = T)
27
28 # Conclusions:
29 # No significant difference in incidence of adverse event between TNFi
30 and BARI
31 # Increased incidence of lack of efficacy for TNFi compared to BARI.
32 # CAREFUL: using a Fine-Gray model allows us to make claim about the
33 association between a covariate and the direction of the increase in
34 incidence, but we can't quantify the magnitude of the increase in
35 incidence.
36
37
38 Clean tables with Hazard ratios with confidence intervals & p-values
39
40 # Adverse event
41 ploufrows <- as.character(summary(est)$term)
42 ploufcols <- c("HR","95%CI","p")
43 coxtable_ae <- matrix(data = NA, nrow = length(ploufrows), ncol =
44 length(ploufcols))
45 rownames(coxtable_ae) <- ploufrows
46 colnames(coxtable_ae) <- ploufcols
47 plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
48
49 for(row in ploufrows)
50 {
51   coxtable_ae[row,"HR"] <- formattable(plouf[term %in% row,
52 estimate])
53   coxtable_ae[row,"95%CI"] <- paste0(formattable(plouf[term %in% row,
54 `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
55   coxtable_ae[row,"p"] <- writepvalue(plouf[term %in% row, p.value])
56 }
57
58
59
60

```

```

1
2
3
4   row.names(coxtable_ae)[1] <- c("TNFi vs BARI Advserere Events")
5
6   # Lack of efficacy
7   ploufrows <- as.character(summary(est2)$term)
8   ploufcols <- c("HR","95%CI","p")
9   coxtable_lae <- matrix(data = NA, nrow = length(ploufrows), ncol =
10  length(ploufcols))
11  rownames(coxtable_lae) <- ploufrows
12  colnames(coxtable_lae) <- ploufcols
13  plouf <- summary(est2, conf.int = T, exponentiate = T) %>% setDT()
14
15
16  for(row in ploufrows)
17  {
18    coxtable_lae[row,"HR"] <- formattable(plouf[term %in% row,
19    estimate])
20    coxtable_lae[row,"95%CI"] <- paste0(formattable(plouf[term %in%
21    row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
22    coxtable_lae[row,"p"] <- writepvalue(plouf[term %in% row,
23    p.value])}
24
25
26  row.names(coxtable_lae)[1] <- c("TNFi vs BARI Lack of Eff")
27
28  # output
29  coxtable_ae
30  coxtable_lae
31
32  write.xlsx(coxtable_ae, file="./3_clean_output/BARI vs TNFi HR
33  competing risk Fine-Gray AE.xlsx") # saving excel file
34  write.xlsx(coxtable_lae, file="./3_clean_output/BARI vs TNFi HR
35  competing risk Fine-Gray LAE.xlsx") # saving excel file
36
37
38  (BARI vs OMA)
39
40  Cumulative incidence function
41  BARI_comp <- copy(BARI_DATA)
42
43  #General
44
45  BARI_comp[stop_reasons == "ADVERSE_EVENT",status := 1]
46  BARI_comp[stop_reasons == "NOT_EFFECTIVE", status := 2]
47  BARI_comp[stop_reasons == "OTHER" | stop_reasons == "REMISSION",
48  status := 3]
49  BARI_comp[stop_reasons == "CONTINUE", status := 0]
50  BARI_comp$cohort <- as.factor(BARI_comp$cohort)
51
52
53  library(reshape)
54
55  BARI_comp_B <- BARI_comp[cohort %in% c("BARI")] #BARI only
56
57
58
59
60

```

```

1
2
3 ci_BARI <- Cuminc(time = "time_on_drug",status = "status", data =
4 BARI_comp_B)
5 ci_BARI <- ci_BARI[, -c(2,6,7,8,9)]
6 ci_long_BARI <- reshape2::melt(ci_BARI,id.vars = "time")
7
8
9 BARI_comp_0 <- BARI_comp[cohort %in% c("OMA")] #OMA only
10 ci_OMA <- Cuminc(time = "time_on_drug",status = "status", data =
11 BARI_comp_0)
12 ci_OMA <- ci_OMA[, -c(2,6,7,8,9)]
13 ci_long_OMA <- reshape2::melt(ci_OMA,id.vars = "time")
14
15 ci_long_BARI$cohort <- 0
16 ci_long_OMA$cohort <- 1
17 ci_long_2 <- rbind(ci_long_BARI,ci_long_OMA)
18 ci_long_2$cohort <- as.factor(ci_long_2$cohort)
19
20
21 plot3 <- ggplot(data = ci_long_2, aes(x = time,
22                                     y = value,
23                                     linetype =
24 interaction(cohort,variable),
25                                                     col =
26 interaction(cohort,variable))) +
27   geom_line(size = 0.75) +
28   scale_color_manual(name = "",
29                      values =
30 c("#08306B", "#08306B", "#238B45", "#238B45", "#FD8D3C", "#FD8D3C"),
31                      labels = c("Adverse Event (BARI)", "Adverse Event
32 (OMA)", "Ineffectiveness (BARI)", "Ineffectiveness (OMA)", "Other
33 (BARI)", "Other (OMA)"))+
34   scale_linetype_manual(name="",
35                         values = c(1,3,1,3,1,3),
36                         labels = c("Adverse Event (BARI)", "Adverse
37 Event (OMA)", "Ineffectiveness (BARI)", "Ineffectiveness (OMA)", "Other
38 (BARI)", "Other (OMA)"))+
39   scale_x_continuous(name = "Time", limits = c(1,365)) +
40   scale_y_continuous(name = "Cumulative incidence", limits =
41 c(0.0,0.3)) +
42   theme_bw()+
43   theme(strip.text.y = element_blank(),
44         strip.background = element_blank(),
45         axis.line.x = element_line(size = 0.5),
46         axis.text = element_text(face = "bold", colour = "black"),
47         legend.position="right", plot.margin =
48 unit(c(1,3,2,1),"lines"))+
49   #ggtitle("Cumulative incidence functions")+
50   theme(plot.title = element_text(hjust = 0.5))
51
52
53 plot3
54
55
56
57
58
59
60

```


Adjusting variables

```
# Covariates of interest for Cox
```

```
covs <-  
c("cohort", "age_base", "bmi_base", "TC_with_csDMARD", "PREDNISON_STEROID"  
  , "CDAI0", "disease_duration_base_years", "smoker_base", "line_of_therapy"  
  , "gender", "seropositivity_base")
```

Cause-specific hazard model

```
# Rappel: imputed_data1 = BARI vs TNFi  
#           imputed_data2 = BARI vs OMA
```

```
# Transition matrix definition
```

```
library(mstate)  
tmat <- trans.comprisk(2, names = c("event-free", "ae", "lae"))  
tmat
```

```
imputed_data2_long <- complete(imputed_data2, action = "long") %>%  
setDT()  
imputed_data2_long[, stop_ae := fifelse(stop_reasons ==  
  "ADVERSE_EVENT", 1, 0)]  
imputed_data2_long[, stop_lae := fifelse(stop_reasons ==  
  "NOT_EFFECTIVE", 1, 0)]  
imputed_data2_long[, stop_other := fifelse(stop_reasons == "OTHER" |  
  stop_reasons == "REMISSION", 1, 0)]  
#[, continue := fifelse(stop_reasons == "CONTINUE", 1, 0)]  
imputed_data2_long[, continue := fifelse(stop_reasons == "OTHER" |  
  stop_reasons == "REMISSION" | stop_reasons == "CONTINUE", 1, 0)]
```

```
M <- imputed_data2$m
```

```
mice_fit <- lapply(1:M, function(m){
```

```
  # subset
```

```
  data_sub <- imputed_data2_long[.imp == m]
```

```
  mst_hosp <- msprep(time =  
c("time_on_drug", "time_on_drug", "time_on_drug"),  
  status = c("continue", "stop_ae", "stop_lae"),  
  data = as.data.frame(data_sub),  
  trans = tmat,  
  keep = covs)
```

```
  # get covariates
```

```
tmp <- expand.covs(mst_hosp, covs, append = TRUE, longnames = T)  
tmp_cov <- grep(paste0(covs, "."), collapse = "|"), names(tmp), value = T)
```

```
  # fit
```

```
coxph(as.formula(paste0("Surv(Tstart, Tstop, status) ~",  
  paste0(tmp_cov, collapse = " + ")),
```

```

1
2
3                                     "+ strata(trans))),
4     data = tmp,
5     method = "breslow")
6
7  }) %>%
8     as.mira()
9
10  est <- pool(mice_fit)
11  summary(est, conf.int = T, exponentiate = T)
12
13
14  # Transition 1 = Adverse Event
15  # Transition 2 = Lack of Efficacy
16
17  # Conclusion
18  # => No difference between OMA and BARI for Adverse Event (ae) and for
19  Lack of Event (lae)
20
21  Cleaner table with Hazard ratios with confidence intervals & p-values
22
23  ploufrows <- as.character(summary(est)$term)
24  ploufcols <- c("HR","95%CI","p")
25  coxtable_csh2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
26  length(ploufcols))
27  rownames(coxtable_csh2) <- ploufrows
28  colnames(coxtable_csh2) <- ploufcols
29  plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
30
31  for(row in ploufrows)
32  {
33    coxtable_csh2[row,"HR"] <- formattable(plouf[term %in% row,
34    estimate])
35    coxtable_csh2[row,"95%CI"] <- paste0(formattable(plouf[term %in%
36    row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
37    coxtable_csh2[row,"p"] <- writepvalue(plouf[term %in% row,
38    p.value])}
39
40
41  row.names(coxtable_csh2)[1:2] <- c("OMA vs BARI Adverse event", "OMA
42  vs BARI Lack of Eff" )
43  coxtable_csh2
44
45  Subdistribution hazard model (Fine-Gray)
46  # Status variable
47  imputed_data2_long[stop_reasons == "ADVERSE_EVENT",status := 1]
48  imputed_data2_long[stop_reasons == "NOT_EFFECTIVE", status := 2]
49  imputed_data2_long[stop_reasons == "OTHER" | stop_reasons ==
50  "REMISSION", status := 3]
51  imputed_data2_long[stop_reasons == "CONTINUE", status := 0]
52
53
54  imputed_data2_long$line_of_therapy <-
55  as.factor(imputed_data2_long$line_of_therapy)
56
57
58
59
60

```

```
1
2
3 imputed_data2_long$seropositivity_base <-
4 as.factor(imputed_data2_long$seropositivity_base)
5 levels(imputed_data2_long$cohort) <- c("0","1")
6 levels(imputed_data2_long$line_of_therapy) <- c("0","1","2","3")
7 levels(imputed_data2_long$gender) <- c("0","1")
8 levels(imputed_data2_long$smoker_base) <- c("2","1","0")
9 levels(imputed_data2_long$smoker_base)
10 levels(imputed_data2_long$seropositivity_base) <- c("0","1")
11
12
13 M <- imputed_data2$m
14
15 # First loop to get estimates for event = 1: ADVERSE EVENT
16
17 mice_fit <- lapply(1:M,function(m){
18   # subset
19   BARI_toto <- imputed_data2_long[.imp == m]
20
21   #subdistribution hazard model
22   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
23 BARI_toto[,..covs],failcode = 1,cencode = 0)
24
25
26 }) %>%
27   as.mira()
28 est <- pool(mice_fit)
29 summary(est, conf.int = T, exponentiate = T)
30
31 # Second loop to get estimates for event = 2: LACK OF EFFICACY
32 mice_fit2 <- lapply(1:M,function(m){
33
34   # subset
35   BARI_toto <- imputed_data2_long[.imp == m]
36
37   # subdistribution hazard model
38   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
39 BARI_toto[,..covs],failcode = 2,cencode = 0)
40
41
42 }) %>%
43   as.mira()
44 est2 <- pool(mice_fit2)
45 summary(est2, conf.int = T, exponentiate = T)
46
47 # Conclusions:
48 # No significant difference in incidence of "adverse event" and "lack
49 of efficacy" between TNFi and BARI
50
51
52 # CAREFUL: using a Fine-Gray model allows us to make claim about the
53 association between a covariate and the direction of the increase in
54 incidence, but we can't quantify the magnitude of the increase in
55 incidence.
56
57
58
59
60
```

Cleaner Table with Hazard ratios with confidence intervals & p-values

```

1
2
3
4
5 # Adverse event
6 ploufrows <- as.character(summary(est)$term)
7 ploufcols <- c("HR", "95%CI", "p")
8 coxtable_ae2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
9 length(ploufcols))
10 rownames(coxtable_ae2) <- ploufrows
11 colnames(coxtable_ae2) <- ploufcols
12 plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
13
14
15 for(row in ploufrows)
16 {
17   coxtable_ae2[row, "HR"] <- formattable(plouf[term %in% row,
18 estimate])
19   coxtable_ae2[row, "95%CI"] <- paste0(formattable(plouf[term %in%
20 row, `2.5 %`]), "-", formattable(plouf[term %in% row, `97.5 %`]))
21   coxtable_ae2[row, "p"] <- writepvalue(plouf[term %in% row,
22 p.value])}
23
24 row.names(coxtable_ae2)[1] <- c("OMA vs BARI Advserere Events")
25
26
27 # Lack of efficacy
28 ploufrows <- as.character(summary(est2)$term)
29 ploufcols <- c("HR", "95%CI", "p")
30 coxtable_lae2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
31 length(ploufcols))
32 rownames(coxtable_lae2) <- ploufrows
33 colnames(coxtable_lae2) <- ploufcols
34 plouf <- summary(est2, conf.int = T, exponentiate = T) %>% setDT()
35
36
37 for(row in ploufrows)
38 {
39   coxtable_lae2[row, "HR"] <- formattable(plouf[term %in% row,
40 estimate])
41   coxtable_lae2[row, "95%CI"] <- paste0(formattable(plouf[term %in%
42 row, `2.5 %`]), "-", formattable(plouf[term %in% row, `97.5 %`]))
43   coxtable_lae2[row, "p"] <- writepvalue(plouf[term %in% row,
44 p.value])}
45
46 row.names(coxtable_lae2)[1] <- c("OMA vs BARI Lack of Eff")
47
48
49 #Output
50 coxtable_ae2
51 coxtable_lae2
52
53 write.xlsx(coxtable_ae2, file="./3_clean_output/BARI vs OMA HR
54 competing risk Fine-Gray AE.xlsx") # saving excel file
55 write.xlsx(coxtable_lae2, file="./3_clean_output/BARI vs OMA HR
56 competing risk Fine-Gray LAE.xlsx") # saving excel file
57
58
59
60

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1. Saving

```
save.image(file="./3_clean_output/full_workspaces/workspace_1.RData")
```

For peer review only

2 - LDA and REM ANALYSIS

10/11/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)

rm(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))
```

```
load("../1_datamanaged_files/datamanaged.Rdata")
```

This code aims at providing estimates for the remission rates of the different treatments groups REM = REMission LDA = Low Disease Activity

Both outcome are base on the CDAI CDAI = Clinical Disease Activity Index

CDAI is an index computed by the physician, which scores the severity of the disease.

1. [0] Exploration

See all available raw CDAI measures :

```
BARI_long[, group := "non-BARI"]
BARI_long[drug == "BIOLOGIC_BARICITINIB", group := "BARI"]
```

```
nrow(BARI_DATA)
summary(BARI_DATA[, .(CDAI0_raw, CDAI12_raw)])
```

1. [1] CARRAC (confirm covariates for confounding and for attrition)

For LDA with updated function

```
library(modules)
source_comp_eff <- modules::use("ETAPE_2_supp_code.R")

LDA_BARI_TNF <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "TNFi")],
  var = "CDAI12",
  thres = 10,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons" ),
  seed = 123)

LDA_BARI_OMA <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "OMA")],
  var = "CDAI12",
  thres = 10,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons"),
  seed = 123)

LDA_BARI_TNF
LDA_BARI_OMA
```

For REM with updated function

```
REM_BARI_TNF <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "TNFi")],
  var = "CDAI12",
  thres = 2.8,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons" ),
  seed = 123)

REM_BARI_OMA <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "OMA")],
```

```
1
2
3   var = "CDAI12",
4   thres = 2.8,
5   ttt_var = "cohort",
6   ref_ttt = "BARI",
7   counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
8                   "line_of_therapy", "CDAI0"),
9   attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
10                "line_of_therapy", "CDAI0", "stop_reasons"),
11   seed = 123)
```

```
13
14 REM_BARI_TNF
15 REM_BARI_OMA
```

16
17 This methods was developed by Mongin et al,
18 <https://ard.bmj.com/content/early/2022/01/12/annrheumdis-2021-221477>

20 Pooled table

```
21 table <- rbind(LDA_BARI_TNF, LDA_BARI_OMA, REM_BARI_TNF, REM_BARI_OMA)
22
23 write.xlsx(table, file = "./3_clean_output/table_LDA_REM_CARRAC.xlsx",
24            row.names = F)
```

27 1. Saving

```
28
29 save.image(file="./3_clean_output/full_workspaces/workspace_2.RData")
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```


2 - LDA and REM supp CODE

10/11/2020

```

1  {r setup, include=FALSE} import("data.table") import("plyr")
2  import("data.table") import("mice") import("ipw") import("survey")
3  import("geepack") import("futile.logger") import("emmeans")
4  import("stats") import("survival")
5
6
7
8
9

```

function to perform checks on data

```

10 ``{r setup, include=FALSE} check_data = function(datain, var = "CDAI_fu", ttt_var = "ttt",
11 ref_ttt = "ttt_ref", ID_ttt = NULL, othervar = c())
12
13 {
14
15 data <- setDT(copy(datain))
16
17 vartochek <- Reduce(union,list(var,ttt_var,othervar)) notindata <-
18 setdiff(vartochek,names(data))
19
20 if(length(notindata)>0){ stop(paste0("the variables",paste0(notindata,collapse = ",")," are
21 not in the dataplease correct")) }
22
23 # force ttt as var name setnames(data,ttt_var,"ttt")
24
25 if( data[,uniqueN(ttt)]>2){ stop("there are more than two treatments. The analysis has
26 been implemented only for 2 treatments") }
27
28 if(!any(data$ttt == ref_ttt)){ stop(paste0("The variable",ttt_var," does not contain any
29 ",ttt_ref," value")) }
30
31 data[,ttt := relevel(as.factor(ttt),ref_ttt)] if(is.null(ID_ttt))
32 { data[,ID_ttt := .I] }else{ setnames(data,ID_ttt,"ID_ttt") data[,N := .N,by = ID_ttt]
33 if(any(data$N>1)){ stop("there are",data[N>1,uniqueN(ID_ttt)]," treatment course which
34 have more than one entry in the table. Each row should be an unique treatment") } }
35 return(data) }
36
37 adjusted_model = function(data, weights = NULL, covariates = NULL){
38
39 # transform char to factor to_fact <- data[,lapply(.SD,class)] %>% transpose(keep.names =
40 "var") %>% .[V1 == "character",var]
41
42 data[,c(to_fact) := lapply(.SD,factor),.SDcols = to_fact]
43
44 #droplevels facto_vars <- data[,lapply(.SD,class)] %>% transpose(keep.names = "var") %>
45 % .[V1 == "factor",var] data[,c(facto_vars) := lapply(.SD,droplevels),.SDcols = facto_vars]
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

1
2
3 # define formula formula <- as.formula(paste0("LDA ~",paste0(c("ttt",covariates),collapse
4 = " + "))
5
6 if(!is.null(covariates)){ # fit fit <- geeglm(formula, data = data, id = ID_ttt, family =
7 gaussian) }else{ fit <- geeglm(LDA ~ ttt, data = data, weights = weights, id = ID_ttt, family =
8 gaussian) }
9
10 fitsummary <- summary(fit) # create table with difference between the two treatments diff
11 <- data.table(ttt = "diff", LDA = fitsummary$coefficients[2,"Estimate"], LDA_var =
12 fitsummary$coefficients[2,"Std.err"]^2, LDA_sup = fitsummary$coefficients[2,"Estimate"] +
13 1.96*fitsummary$coefficients[2,"Std.err"], LDA_inf = fitsummary$coefficients[2,"Estimate"]
14 - 1.96*fitsummary$coefficients[2,"Std.err"], methods = "CC_adjusted")
15
16 # marginal effects: margi_df <- emmeans(fit, "ttt") %>% as.data.table()
17
18 margi_df[,methods := "CC_adjusted"] setnames(margi_df,"emmean","LDA")
19 margi_df[,LDA_inf := LDA - 1.96*SE] margi_df[,LDA_sup := LDA + 1.96*SE]
20 margi_df[,LDA_var := SE^2]
21
22 output <- rbind(diff,margi_df[,.(ttt,LDA,LDA_sup,LDA_inf,LDA_var,methods)])
23
24 return(list(output = output,fit = fit)) }
25
26
27
28
29
30
31 # Not adjusted complete case imputation
32
33
34
35 ```{r setup, include=FALSE}
36
37 export("CC_raw")
38 CC_raw <- function(datain,
39 # data
40 var = "CDAI_fu",
41 # variable measuring effectiveness
42 thres = 10,
43 # threshold for remission or LDA
44 ttt_var = "ttt",
45 ref_ttt = "ttt_ref")
46 # variable name containing the treatment
47 {
48 data <- check_data(datain,var,ttt,ref_ttt)
49 # raw proportion
50 raw_prop <- data[!is.na(get(var)),
51 .(LDA = sum(get(var)<=thres)/.N,
52 methods = "CC_raw",
53 N = .N),
54 by = ttt]
55
56
57
58
59
60

```

```

1
2
3   # calculation of the Standard error
4   raw_prop[,c("LDA_inf","LDA_sup") := lapply(c(-1.96,1.96),function(z)
5   {
6     LDA + z*sqrt(LDA*(1-LDA)/N)
7   }))]
8
9   # difference between treatments
10  diff_tmp <- raw_prop[,.(ttt = "diff",
11                          LDA = LDA[ttt == "ttt_1"]-LDA[ttt ==
12 "ttt_ref" ],
13                          methods = methods[1] ,
14                          SE = (sum(1/N))/2 + 1.96*sqrt(sum( LDA*(1-LDA)/N
15 )))]]
16
17
18  diff_tmp[,LDA_inf := LDA - SE]
19  diff_tmp[,LDA_sup := LDA + SE]
20
21  # bind outputs
22  output <- rbind(diff_tmp[,.(ttt,LDA,LDA_inf,LDA_sup,methods)],
23                 raw_prop[,-"N"])
24
25  # change name back
26  setnames(output,"ttt",ttt_var)
27  return(output)
28 }
29
30

```

Adjusted complete case imputation

```

33 ```{r setup, include=FALSE} export("CC_adjusted") CC_adjusted = function(datain, var =
34 "CDAI_fu", thres = 10, ttt_var = "ttt", ref_ttt = "ttt_ref", covariates =
35 c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAIO") ) # variable
36 name containing the treatment { data <- check_data(datain,var,ttt_var,ref_ttt) data[,LDA :=
37 get(var) <= thres] output <- adjusted_model(data = data[!is.na(get(var))], covariates =
38 covariates)$output
39
40
41 output[,methods := "CC_adjusted"] # change name back setnames(output,"ttt",ttt_var)
42 return(output) }
43
44
45

```

LOCF imputation

```

46
47
48
49
50 ```{r setup, include=FALSE}
51 export("LOCF")
52
53 LOCF <- function(datain,
54                  var = "CDAI_fu",
55                  var_before = "CDAI_beforefu",
56
57
58
59

```

```

1
2
3         thres = 10,
4         ttt_var = "ttt",
5         ref_ttt = "ttt_ref",
6         covariates =
7 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0")
8 ) {
9   data <- copy(datain)
10
11   data <- check_data(datain, var, ttt_var, ref_ttt)
12   data[is.na(get(var)), c(var) := get(var_before)]
13   data[,LDA := get(var) <= thres]
14
15
16   output <- adjusted_model(data = data,
17                             covariates = covariates)$output
18
19   output[,methods := "LOCF"]
20   # change name back
21   setnames(output, "ttt", ttt_var)
22   return(output)
23 }
24
25
26

```

Lundex imputation

```

27
28
29 ``{r setup, include=FALSE} export("Lundex") Lundex <- function(datain, var = "CDAI_fu",
30 thres = 10, ttt_var = "ttt", ref_ttt = "ttt_ref", treatment_duration = "treatment_duration",
31 stop_var = "stopany", covariates =
32 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0"), boot_num =
33 1000) {
34
35   data <- check_data(datain, var, ttt_var, ref_ttt) data[,LDA := get(var) <= thres] ####
36   bootstrap for SE data[,tmp := 1] # replicated data for bootstrap replicateddata <-
37   data[CJ(tmp = 1, boot = 1:boot_num), on = "tmp", allow.cartesian=TRUE] # sample with
38   replacement for each boot sampled_idx <- replicateddata[,I[sample(1:N, replace = T)], by =
39   boot]$V1 bootstrapdata <- replicateddata[sampled_idx]
40
41
42   # raw proportions raw_prop <- bootstrapdata[!is.na(get(var)), { adjusted_model(data
43   = .SD)$output %>% .[ttt != "diff", (ttt, LDA_raw = LDA) ] }, by = .(boot)]
44
45   # surv analysis for each bootstrapped dataset surv_formula <-
46   as.formula(paste0("Surv(", treatment_duration, ", ", stop_var, ") ~ ttt"))
47
48   surv_coeff <- bootstrapdata[, { temp.km <- survfit(surv_formula, data = .SD) list(surv =
49   summary(temp.km, times = 1)$surv, ttt = gsub("ttt=", "", unique(summary(temp.km)
50   $strata))) }, by = boot]
51
52
53   # LDA: LDA raw * surv coeff tmp_bootstrap <- merge(raw_prop, surv_coeff, by =
54   c("boot", "ttt")) tmp_bootstrap[,LDA := LDA_raw*surv]
55
56
57
58
59
60

```

```

1
2
3 # difference between treatments diff_boot <- tmp_bootstrap[,(ttt = "diff", LDA = LDA[ttt !=
4 ref_ttt] - LDA[ttt == ref_ttt]), by = boot]
5
6 tot_bottstrap <- rbind(diff_boot[,.(ttt, LDA, boot)], tmp_bootstrap[,.(ttt, LDA, boot)])
7
8 # calculate the mean and the SE: output <- tot_bottstrap[,.(LDA = mean(LDA), LDA_sup =
9 quantile(LDA, 0.975), LDA_inf = quantile(LDA, 0.025) ), by = ttt] # change name back
10 output[,methods := "LUNDEX"]
11
12 setnames(output,"ttt",ttt_var) return(output) }
13
14
15
16

```

```

17 # non-responder imputation
18
19

```

```

20 ```{r setup, include=FALSE}
21 export("NRI")
22 NRI = function(datain,
23               var="CDAI_fu",
24               thres = 10,
25               ttt_var = "ttt",
26               ref_ttt = "ttt_ref",
27               covariates =
28 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0")
29 )
30 # variable name containing the treatment
31 {
32   data <- check_data(datain, var, ttt_var, ref_ttt)
33   data[,LDA := get(var) <= thres]
34   data[is.na(LDA),LDA := 0] # missing are non responders
35   output <- adjusted_model(data = data,
36                             covariates = covariates)$output
37
38   # change name back
39   setnames(output, "ttt", ttt_var)
40   output[,methods := "NRI"]
41   return(output)
42 }
43
44
45
46

```

Inverse probability weighting imputation

```

47
48
49 ```{r setup, include=FALSE} export("IPW") IPW <- function(datain, var = "CDAI_fu", thres =
50 10, ttt_var = "ttt", ref_ttt = "ttt_ref", counfounders =
51 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0"), attrition =
52 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0", "stopreason")) {
53
54   data <- check_data(datain, var, ttt_var, ref_ttt, othervar = c(counfounders, attrition))
55
56
57
58
59
60

```

```

1
2
3 data[,ttt2 := as.numeric(ttt != ref_ttt)] # weight for confounding formula_coeff <-
4 paste0("~",paste0(counfunders,collapse = "+")) function_call <- paste0('IPWT <-
5 ipwpoint( exposure = ttt2 , family = "binomial", link = "logit", numerator = ~ 1,
6 denominator =',formula_coeff,', data = data, trunc = 0.01 )') eval(parse(text = function_call))
7 datasw<- IPW Tipw.weights
8
9
10 # weights for attrition formula_attr <- paste0("~",paste0(attrition,collapse = "+"))
11 data[,MISS := as.numeric(is.na(get(var)))] function_call <- paste0('IPCT <-
12 ipwpoint( exposure = MISS , family = "binomial", link = "logit", numerator = ~ 1,
13 denominator =',formula_attr,', data = data )') eval(parse(text = function_call)) data
14 swc<- IPC Tipw.weights
15
16 dataNoNA <- na.omit(data[,.(ttt,get(var),sw,swc,ID_ttt) %>%
17 setNames(c("ttt",var,"sw","swc","ID_ttt"))]) dataNoNA[,LDA := as.numeric(get(var) <=
18 thres)]
19
20
21 output <- adjusted_model(data = dataNoNA, weights = dataNoNASw*dataNoNASwc)
22 $output
23
24 output[,methods := "IPW"]
25
26 # change name back setnames(output,"ttt",ttt_var) return(output)
27
28 }
29
30
31
32 # Confounder-Adjusted Response Rate with Attrition Correction (CARRAC)
33 imputation
34
35
36 ```{r setup, include=FALSE}
37 export("CARRAC")
38 CARRAC <- function(datain,
39                     var = "CDAI_fu",
40                     thres = 10,
41                     ttt_var = "ttt",
42                     ref_ttt = "ttt_ref",
43                     counfunders =
44                     c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAI0"),
45                     attrition =
46                     c("Disease_duration","concomitantCsDMARD",
47                       "Prev_bDMARD3","CDAI0","stopreason"),
48                     seed = NA) {
49
50   data <- check_data(datain,var,ttt_var,ref_ttt)
51   dataS <- data[,.SD,.SDcols =
52   c("ID_ttt",var,"ttt",union(counfunders,attrition))]
53
54
55
56
57
58
59
60

```

```

1
2
3   impute_data <- mice(
4     dataS,
5     m = 10,
6     method = "pmm",
7     maxit = 5,
8     printFlag = F, seed = seed
9   )
10  # open the data
11  impute_data_complete <- setDT(complete(impute_data, action = "long"))
12  # calculate LDA
13  impute_data_complete[, LDA := get(var) <= thres]
14
15
16  # get LDA and error for each imputation
17  res_mice <- lapply(seq(1:impute_data$m), function(imp){
18
19    adjusted_model(data = impute_data_complete[.imp == imp],
20                  covariates = counfunders)$output
21
22  }) %>% rbindlist()
23
24  res_mice_2 <- lapply(seq(1:impute_data$m), function(imp){
25
26    adjusted_model(data = impute_data_complete[.imp == imp],
27                  covariates = counfunders)$fit
28
29  })
30
31
32  test <- pool(res_mice_2)
33  df_pval <- summary(test) %>% as.data.table()
34  p.output <- df_pval[grepl("ttt", term), p.value]
35
36  # pooling
37  pool_res <- res_mice[,.(
38    LDA_mi = mean(LDA),
39    w = mean(LDA_var),
40    m = .N,
41    b = 1/(.N-1)*sum( (LDA-mean(LDA))^2 )
42    ), by = ttt]
43
44
45  pool_res[,LDA_var := w + (1+1/m)*b]
46  pool_res[,LDA_sd := sqrt(LDA_var)]
47
48  # mean, 95% CI
49  output <- pool_res[,.(ttt,
50                        LDA_mi,
51                        LDA_mi + 1.96*LDA_sd,
52                        LDA_mi-1.96*LDA_sd) %>%
53                        setNames(c("ttt", "LDA", "LDA_sup", "LDA_inf"))]
54
55  output[, methods := "CARRAC"]
56
57
58
59
60

```

```
1
2
3     output[ttt == "diff",p := p.output]
4
5     # change name back
6     setnames(output,"ttt",ttt_var)
7
8     return(output)
9 }
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

For peer review only

3 - FINAL FIGURES CODE

10/11/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)
library(patchwork) # package to compose multiplots !
library(ggpubr)
library(grid)

rm(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext())$path)

load("./3_clean_output/full_workspaces/workspace_1.RData")
load("./3_clean_output/full_workspaces/workspace_2.RData")
load("./3_clean_output/full_workspaces/workspace_3.RData")
```

1. Common theme

```
theme_benoit = function(){
  theme_pubclean()+
    theme(panel.grid.major.x = element_line(linetype = "dotted", colour
= "grey50"),
          panel.grid.major.y = element_line(linetype = "dotted", colour
= "grey50"),
          axis.title.y = element_text(margin = margin(r = .2, unit =
"cm")),
          axis.title.x = element_text(margin = margin(t = .2, unit =
```

```
1  
2  
3 "cm")),  
4 plot.title = element_text(margin = margin(b = .5, unit =  
5 "cm")))  
6 }  
7  
8  
9
```

1. [0] Mini explanation

TC lenght

```
13 BARI_DATA[,time_on_drug_year := time_on_drug/365.25]  
15  
16 p1 <- ggplot(BARI_DATA)+  
17   geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =  
18   1/12)+  
19  
20   scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+  
21   labs(x = "Duration of observation (years)",  
22        y = "Number of TC",  
23        title = "Time of observation for all included TC")+  
24   ylim(-25,NA)+  
25   theme_benoit()  
26 p1  
27  
28  
29 p2 <- ggplot(BARI_DATA)+  
30   geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =  
31   "grey80")+  
32   theme_void()  
33  
34 plot_mini_exploration <- p1 + inset_element(p2,0.01,0.05,0.99,0.2)  
35 plot_mini_exploration  
36
```

Saving plot

```
39 png("./3_clean_output/figures/PLOT_Exploration_TC_duration.png",  
40     width = 7,  
41     height = 5,  
42     units = "in",  
43     res = 300) # opening graphic device  
44 plot_mini_exploration  
45 dev.off() # closing graphic device  
46
```

TC lenght for BARI only

```
49 data_sub <- BARI_DATA[cohort == "BARI"]  
50  
51 p1 <- ggplot(data_sub)+  
52   geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =  
53   1/13, fill = "red3")+  
54  
55   scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+  
56  
57  
58  
59
```

```
1
2
3     labs(x = "Duration of observation (years)",
4           y = "Number of TC",
5           title = "A - BARI")+
6     ylim(-11,50)+
7     theme_benoit()
8 p1
9
10
11 p2 <- ggplot(data_sub)+
12     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
13 "grey80")+
14     theme_void()
15
16 plot_mini_exploration_bari <- p1 +
17 inset_element(p2,0.01,0.05,0.99,0.2)
18 plot_mini_exploration_bari
19
20 TC lenght for TNFi only
21
22 data_sub <- BARI_DATA[cohort == "TNFi"]
23
24 p1 <- ggplot(data_sub)+
25     geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
26 1/13, fill = "green2")+
27
28     scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
29     labs(x = "Duration of observation (years)",
30          y = "Number of TC",
31          title = "B - TNFi")+
32     ylim(-11,50)+
33     theme_benoit()
34 p1
35
36
37 p2 <- ggplot(data_sub)+
38     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
39 "grey80")+
40     theme_void()
41
42 plot_mini_exploration_tnfi <- p1 +
43 inset_element(p2,0.01,0.05,0.99,0.2)
44 plot_mini_exploration_tnfi
45
46 TC lenght for OMA only
47
48 data_sub <- BARI_DATA[cohort == "OMA"]
49
50
51 p1 <- ggplot(data_sub)+
52     geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
53 1/13, fill = "blue2")+
54
55     scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
56
57
58
59
60
```

```

1
2
3     labs(x = "Duration of observation (years)",
4           y = "Number of TC",
5           title = "C - OMA")+
6     ylim(-11,50)+
7     theme_benoit()
8 p1
9
10
11 p2 <- ggplot(data_sub)+
12     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
13 "grey80")+
14     theme_void()
15
16 plot_mini_exploration_oma <- p1 + inset_element(p2,0.01,0.05,0.99,0.2)
17 plot_mini_exploration_oma
18
19 multiplot
20
21 multi_plot <- plot_mini_exploration_bari + plot_mini_exploration_tnfi
22 + plot_mini_exploration_oma
23 multi_plot
24
25 median(BARI_DATA[cohort == "BARI", time_on_drug])
26 median(BARI_DATA[cohort == "TNFi", time_on_drug])
27 median(BARI_DATA[cohort == "OMA", time_on_drug])
28
29 Saving plot
30
31 png("./3_clean_output/figures/
32 PLOT_Exploration_TC_duration_3_groups.png",
33     width = 9,
34     height = 5,
35     units = "in",
36     res = 300) # opening graphic device
37 multi_plot
38 dev.off() # closing graphic device
39
40
41
42
43
44

```

1. [1] Survival analysis

Forest plot BARI vs TNFi + BARI vs OMA

```

45 meanall <- summary(BARI1.adj.mi)$coefficients[1:14,"exp(coef)"]
46 lowerall <- summary(BARI1.adj.mi)$conf.int[1:14,"lower .95"]
47 upperall <- summary(BARI1.adj.mi)$conf.int[1:14,"upper .95"]
48 textall <- c("Treatment (vs BARI)", "Age (decades)", "BMI",
49 "Concomitant csDMARD", "Concomitant glucocorticoid", "CDAI score (10
50 pts)", "Disease duration (decades)", "Current smoker (vs non-smoker)",
51 "Former smoker (vs non-smoker)", "2nd line therapy (vs 1st)", "3rd
52 line therapy (vs 1st)", "4th or later line (vs 1st)", "Female gender",
53 "Seropositivity (RF or ACPA)")
54
55
56
57
58
59

```

```

1
2
3 dfall1 <- data.table(textall, meanall, lowerall, upperall)
4 dfall1[,ttt := "TNFi"]
5
6 meanall <- summary(BARI2.adj.mi)$coefficients[1:14,"exp(coef)"]
7 lowerall <- summary(BARI2.adj.mi)$conf.int[1:14,"lower .95"]
8 upperall <- summary(BARI2.adj.mi)$conf.int[1:14,"upper .95"]
9 dfall2 <- data.table(textall, meanall, lowerall, upperall)
10 dfall2[, ttt := "OMA"]
11
12
13 dfall <- rbind(dfall1,dfall2)
14 dfall$textall <- factor(dfall$textall, levels = rev(textall))
15
16 text_high <- textGrob("\u2192 Reduces \ndrug maintenance",
17 gp=gpar(fontsize=8, fontface="bold"))
18 text_low <- textGrob("\u2190 Improves \ndrug maintenance",
19 gp=gpar(fontsize=8, fontface="bold"))
20
21
22 HR_plot <- ggplot(data=dfall,
23 aes(x = textall,
24 y = meanall,
25 ymin = lowerall,
26 ymax = upperall,
27 color = ttt))+
28 geom_hline(yintercept =1, linetype=2)+
29 geom_point(size=2,position = position_dodge(width = .7))+
30 geom_errorbar(position = position_dodge(width = .7))+
31 labs(x = "",y = "",color = "")+
32 scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
33 +
34 theme(axis.line.x = element_line(size = 0.5),
35 axis.text = element_text(face = "bold", color = "black"),
36 legend.position="top",
37 legend.key = element_blank(),
38 plot.margin = unit(c(1,3,2,1),"lines"))+
39 coord_flip(clip = "off")+
40 annotation_custom(text_high,
41 xmin=-0.64,xmax=-0.64,ymin=.2,ymax=.2)+
42 annotation_custom(text_low,
43 xmin=-0.64,xmax=-0.64,ymin=-.15,ymax=-.15)+
44 theme_pubclean()+
45 scale_color_manual(breaks = c("OMA","TNFi"),values =
46 c("blue3","green3"),labels = c("OMA","TNFi"))
47
48

```

HR_plot

Saving the plot in PNG file

```
png("./3_clean_output/figures/PLOT FOREST BARI vs TNFi vs OMA HR.png",
```

```
width = 7,
```

```
1
2
3     height = 5.5,
4     units = "in",
5     res = 300)
6
7 HR_plot
8
9
10 dev.off() # closing graphic device
11
12 BARI vs TNFi
13
14 Non-adjusted Kaplan-Meier curves
15
16 BARI vs TNFi
17
18 BARI1[,time_on_drug_year := time_on_drug/365.25]
19
20 surv_object1 <- Surv(time = BARI1$time_on_drug_year, event =
21 BARI1$stop_DMARD) # indicate time on drug and stop variable
22 fit1 <- survfit(surv_object1 ~ cohort, data = BARI1)
23
24 survplot_1 <- ggsurvplot(fit1, data = BARI1, # plot
25                          pval = T,
26                          pval.method = TRUE,
27                          legend.title = "Groups :",
28                          legend.labs = c("BARI", "TNFi"),
29                          xlab = "Time (years)",
30                          xlim = c(0, 2.5),
31                          censor = FALSE,
32                          title = "Non-adjusted model of drug
33 discontinuation \nby type of treatment",
34                          surv.median.line = "v",
35                          linetype = 1,
36                          size = 1.5,
37                          #palette = c("grey78", "grey10"),
38                          palette = c("red3", "green2"), # pour mettre
39 les couleurs
40
41                          ggtheme = theme_benoit(),
42                          risk.table = T)
43
44
45 values <- summary(fit1)$table[,"median"]
46 df <- data.frame(y = .1,x = values+.2,label =
47 as.character(round(values,2)))
48
49 survplot_1$plot <- survplot_1$plot +
50   geom_text(data = df,aes(x,y,label = label), color = c("red3",
51 "green2"), size = 5)
52
53
54
55 print(survplot_1)
56
57
58
59
60
```

1
2
3 Saving surplot

4
5 png("./3_clean_output/figures/PLOT BARI vs TNFi curves non adjusted
6 COLOR.png",
7 width = 7,
8 height = 7, units = "in",
9 res = 300) # opening graphic device
10 survplot_1

11
12 dev.off() # closing graphic device

13
14 Saving the plot curv object for Lilly

15
16 plot_BARI_vs_TNFi_data <- survplot_1\$data.survplot
17 write.xlsx(plot_BARI_vs_TNFi_data, file =
18 "./3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_non_adjust
19 ed.xlsx", row.names = F)

20
21 **Home-made attempt to obtain adjusted curves based on imputed data**

22 dummy_cox_impute1 <- mice::complete(imputed_data1, "long", include =
23 T)

24 dummy_cox_impute1 <- dummy_cox_impute1[dummy_cox_impute1\$.imp != 0,]

25 dummy_cox_impute1\$time_on_drug_year <-
26 dummy_cox_impute1\$time_on_drug/365.25

27
28
29 BARI_fit1 <- survfit(coxph(Surv(time = time_on_drug_year, event =
30 stop_DMARD) ~ cohort+

31 I(age_base/10)+

32 bmi_base+

33 TC_with_csDMARD+

34 PREDNISON_STEROID+

35 CDAI0+

36 I(disease_duration_base_years/10)+

37 C(smoker_base, base=3)+

38 line_of_therapy+

39 gender+

40 seropositivity_base+

41 cluster(patient_id)+

42 strata(cohort), dummy_cox_impute1), data =

43
44 dummy_cox_impute1)

45
46
47
48 survplot_1_adj <- ggsurvplot(BARI_fit1, data = dummy_cox_impute1,
49 variable = "cohort",

50 xlab = "Time (years)",

51 title = "A - BARI vs TNFi",

52 legend.title = "Groups :",

53 legend.labs = c("BARI", "TNFi"),

54 censor = FALSE,

55 xlim = c(0, 2.5),

```

1
2
3
4     surv.median.line = "v",
5     linetype = 1,
6     size = 1.5,
7     ggtheme = theme_benoit(),
8     # palette = c("grey78", "grey10")
9     palette = c("red2", "green3" )+
10    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,\n
11    concomitant glucocorticoid, baseline CDAI, disease duration
12    (decades),\n smoking status, line of therapy, gender, seropositivity")
13
14 # adding days label
15 values <- summary(BARI_fit1)$table[,"median"]
16 df <- data.frame(y = .1,x = values+.1,label =
17 as.character(paste(round(values*365.25,2), "\n days")))
18 df[1,2] <- 1.82
19
20 survplot_1_adj$plot$labels$y <- "Proportion still on drug" # to change
21 the label
22
23 survplot_1_adj$plot <- survplot_1_adj$plot +
24   geom_text(data = df,aes(x,y,label = label), color = c("red3",
25 "green3"), size = 5)
26
27 # adding HR et p val label
28 HR <- data.frame(y = 0.1, x = 0.5, label = paste("HR =",
29 round(exp(BARI1.adj.mi$coefficients[1]), 2), "\n", "p =",
30 round(summary(BARI1.adj.mi)$coefficients[1,"Pr(>|z|)"], 4)  ) )
31
32 survplot_1_adj$plot <- survplot_1_adj$plot +
33   geom_text(data = HR,aes(x,y,label = label) , size = 5)
34
35 # final print
36 survplot_1_adj
37
38 Saving the survival plot in PNG file
39
40 png("./3_clean_output/figures/PLOT BARI vs TNFi curves adjusted.png",
41
42     width = 7,
43     height = 5,
44     units = "in",
45     res = 300)
46
47 survplot_1_adj
48
49 dev.off() # closing graphic device
50
51
52
53

```

Sensitivity analysis (RiskRegression Package)

First plot to get the difference in average treatment effect in percentage


```
1
2
3 dt.out$time_years <- dt.out$time/365.25
4
5 plot.ate.diff <- ggplot(dt.out[type == "meanRisk"], aes(x =
6 time_years, group = level))+
7   geom_vline(xintercept = 1, linetype = 2, size = 1, color = "grey20")
8 +
9   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
10 0.4)+
11   geom_line(aes(y = estimate, color = level), size = 1)+
12
13   scale_colour_manual(values = c("red2","green3"))+
14   scale_fill_manual(values = c("red2","green3"))+
15   theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
16 legend.position="top" )+
17   scale_x_continuous(breaks=seq(0,2.5,0.25)) +
18   scale_y_continuous(labels = scales::percent, limits = c(0,0.65))+
19
20   xlab("Years since intiation of treatment")+
21   ylab("Absolute Risk of treatment discontinuation (%)")+
22   labs(colour="Groups:", fill = "Groups:", title = "A - BARI vs TNFi")
23 +
24   labs(group = "Groups:")
25 +
26   theme_benoit()+
27   theme(axis.title.x = element_text(margin = margin(t = .3,unit =
28 "cm")),
29 axis.title.y = element_text(margin = margin(r = .3,unit =
30 "cm")))
31
32
33 plot.ate.diff
34
35 Saving Plot
36
37 png("./3_clean_output/figures/PLOT BARI vs TNFi curves AIPTW.png",
38 width = 1300, height = 650, res = 120) # opening graphic device
39
40 plot.ate.diff
41
42 dev.off() # closing graphic device
43
44 Second plot to get the ratio in average treatment effect
45
46 plot.ate.ratio <- ggplot(dt.out[type == "ratioRisk"], aes(x = time,
47 group = level))+
48   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
49 0.4)+
50   geom_line(aes(y = estimate, color = level), size = 2)+
51
52   theme_benoit()+
53   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
54   scale_x_continuous(breaks=seq(0,500,50))+
55
56
57
58
59
```

```
1
2
3     scale_y_continuous(limits = c(0.9,4.5))+
4
5     xlab("Days since intiation of treatment")+
6     ylab("Ratio in Average Treatment Effect")+
7     labs(colour="treatment", fill = "treatment")
8
```

```
9
10 plot.ate.ratio
```

11 BARI vs OMA

12 Non-adjusted Kaplan-Meier curves

13 BARI vs OMA

```
14 BARI2[,time_on_drug_year := time_on_drug/365.25]
```

```
15
16 surv_object2 <- Surv(time = BARI2$time_on_drug_year, event =
17 BARI2$stop_DMARD)
18 fit2 <- survfit(surv_object2 ~ cohort, data = BARI2) # this function
19 creates the data for Kaplan Meyer
20 survplot_2 <- ggsurvplot(fit2, data = BARI2, # plot
21     pval = T,
22     pval.method = TRUE,
23     legend.title = "Groups :",
24     legend.labs = c("BARI", "OMA"),
25     xlab = "Time (days)",
26     xlim = c(0, 2.5),
27     censor = FALSE,
28     title = "Non-adjusted model of drug discontinuation by type
29 of treatment",
30     surv.median.line = "v",
31     linetype = 1,
32     size = 1.5,
33     ggtheme = theme_benoit(),
34     #palette = c("grey78", "grey50"),
35     palette = c("red3", "blue2"), # to put colors
36     risk.table = T)
37
```

```
38 survplot_2
```

39 Saving surplot

```
40 png("./3_clean_output/PLOT BARI vs OMA curves non adjusted COLOR.png",
41 width = 1000, height = 600, res = 100) # opening graphic device
```

```
42 survplot_2
```

```
43 dev.off() # closing graphic device
```

44 Saving the plot curv object for Lilly

```

1
2
3 plot_BARI_vs_OMA_data <- survplot_2$data.survplot
4 write.xlsx(plot_BARI_vs_OMA_data, file =
5 ".\3_clean_output\Lilly_curves_excel\plot_BARI_vs_OMA_data_non_adjuste
6 d.xlsx", row.names = F)
7

```

Home-made attempt to obtain adjusted cuves based on imputed data :

```

8
9 dummy_cox_impute2 <- mice::complete(imputed_data2, "long", include =
10 T)
11 dummy_cox_impute2 <- dummy_cox_impute2[dummy_cox_impute2$.imp != 0,]
12 dummy_cox_impute2$time_on_drug_year <-
13 dummy_cox_impute2$time_on_drug/365.25
14
15
16 BARI_fit2 <- survfit(coxph(Surv(time = time_on_drug_year, event =
17 stop_DMARD) ~ cohort+
18
19 I(age_base/10)+
20 bmi_base+
21 TC_with_csDMARD+
22 PRÉDNISON_STEROID+
23 CDAI0+
24 I(disease_duration_base_years/10)+
25 C(smoker_base, base=3)+
26 line_of_therapy+
27 gender+
28 seropositivity_base+
29 cluster(patient_id)+
30 strata(cohort),dummy_cox_impute2), data =
31 dummy_cox_impute2)
32
33 survplot_2_adj <- ggsurvplot(BARI_fit2, data = dummy_cox_impute2,
34 variable = "cohort",
35 xlab = "Time (years)",
36 title = "B - BARI vs OMA",
37 legend.title = "Groups :",
38 legend.labs = c("BARI", "OMA"),
39 censor = FALSE,
40 xlim = c(0, 2.5),
41 surv.median.line = "v",
42 linetype = 1,
43 size = 1.5,
44 ggtheme = theme_benoit(),
45 # palette = c("grey78", "grey50")
46 palette = c("red2", "blue3") # to change colors
47 )+
48
49 labs(caption = "Adjusted for : age, BMI, concomitant csDMARD, \n
50 concomitant glucocorticoid, baseline CDAI, disease duration
51 (decades),\n smoking status, line of therapy, gender, seropositivity")
52
53 # adding Days label
54 values <- summary(BARI_fit2)$table[,"median"]
55 df <- data.frame(y = .1,x = values+.1,label =
56
57
58
59
60

```

```
1
2
3 as.character(paste(round(values*365.25,2), "\n days"))
4 df[1,2] <- 1.82
5
6 survplot_2_adj$plot$labels$y <- "Proportion still on drug" # to change
7 the label
8
9
10 survplot_2_adj$plot <- survplot_2_adj$plot +
11   geom_text(data = df,aes(x,y,label = label), color = c("red3",
12 "blue2"), size = 5)
13
14 # adding HR et pval label
15 HR <- data.frame(y = 0.1, x = 0.5, label = paste("HR =",
16 round(exp(BARI2.adj.mi$coefficients[1]), 2), "\n", "p =",
17 round(summary(BARI2.adj.mi)$coefficients[1,"Pr(>|z|)"], 4) ) )
18
19 survplot_2_adj$plot <- survplot_2_adj$plot +
20   geom_text(data = HR,aes(x,y,label = label) , size = 5)
21
22
23
24
25 # final print
26 survplot_2_adj
27
28 summary(BARI_fit2, times = 1) # to see detailed surv probabilities at
29 given timepoints
30
31 Saving the survival plot in PNG file
32
33 png("./3_clean_output/PLOT BARI vs OMA curves adjusted.png", width =
34 1000, height = 600, res = 100) # opening graphic device
35
36 survplot_2_adj
37
38 dev.off() # closing graphic device
39
40
41 Sensitivity analysis (RiskRegression package)
42
43 First plot to get the difference in average treatment effect in percentage
44
45 dt.out2$time_years <- dt.out2$time/365.25
46
47 plot.ate.diff2 <- ggplot(dt.out2[type == "meanRisk"], aes(x =
48 time_years, group = level))+
49   geom_vline(xintercept = 1, linetype = 2, size = 1, color = "grey20")
50 +
51   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
52 0.3)+
53   geom_line(aes(y = estimate, color = level), size = 1)+
54
55 theme_benoit() + theme(legend.spacing.x = unit(0.2, 'cm'),
56
57
58
59
```

```

1
2
3 legend.position="top" )+
4   scale_x_continuous(breaks=seq(0,2.5,0.25)) +
5   scale_y_continuous(labels = scales::percent, limits = c(0,0.65))+
6
7   xlab("Years since initiation of treatment")+
8   ylab("Absolute Risk of treatment discontinuation (%)")+
9   labs(colour="Groups:", fill = "Groups:", title = "B - BARI vs OMA")+
10  labs(group = "Groups:")+
11
12
13   scale_colour_manual(values = c("red2","blue3"))+
14   scale_fill_manual(values = c("red2","blue3"))

```

```
15
16 plot.ate.diff2
```

```
17
18 Saving Plot
```

```
19
20 png("./3_clean_output/PLOT BARI vs OMA curves AIPTW.png", width =
21 1300, height = 650, res = 120) # opening graphic device
22
```

```
23 plot.ate.diff2
```

```
24
25 dev.off() # closing graphic device
26
```

```
27 Second plot to get the ratio in average treatment effect
28
```

```
29 plot.ate.ratio2 <- ggplot(dt.out2[type == "ratioRisk"], aes(x = time,
30 group = level))+
31   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
32 0.3)+
33   geom_line(aes(y = estimate, color = level), size = 1)+
34
35   theme_benoit()+
36   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
37   scale_x_continuous(breaks=seq(100,400,50))+
38   scale_y_continuous(limits = c(0.8,3))+
39
40
41   xlab("Days since initiation of treatment")+
42   ylab("Ratio in Average Treatment Effect")+
43   labs(colour="treatment", fill = "treatment")
44
```

```
45 plot.ate.ratio2
46
```

47 Multipanel plots

48 To update using patchwork

49 For the paper Non adjusted curves

50 # Creating list object

```
51
52 plots <- list()
53
54
55
56
57
58
59
60
```

```
1
2
3 plots[[1]] <- survplot_1
4 plots[[2]] <- survplot_2
5
6 # Nice function
7
8 multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
9
10 Option 2 putting all data on one panel Kaplan Meier
11
12 BARI vs TNFi vs OMA
13
14 BARI_DATA[,time_on_drug_year := time_on_drug/365.25]
15
16 surv_object3 <- Surv(time = BARI_DATA$time_on_drug_year, event =
17 BARI_DATA$stop_DMARD)
18 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_DATA) # this
19 function creates the data for Kaplan Meyer
20 survplot_3 <- ggsurvplot(fit3, data = BARI_DATA, # plot
21 pval = F,
22 pval.method = F,
23 legend.title = "Groups",
24 legend.labs = c("BARI", "TNFi", "OMA"),
25 xlab = "Time (years)",
26 xlim = c(0, 2.5),
27 censor = FALSE,
28 # title = "Non-adjusted drug discontinuation by type of
29 treatment (Kaplan-Meier)",
30 surv.median.line = "v",
31 linetype = 1,
32 size = 1.5,
33 ggtheme = theme_benoit(),
34 palette = c("red3", "green2", "blue2"), # to put colors
35 risk.table = T)
36
37
38
39 values <- summary(fit3)$table[, "median"]
40 df <- data.frame(y = .1, x = values+.1, label =
41 as.character(paste(round(values*365.25,2), "\n days")))
42 df[3,2] <- 1.72
43
44 survplot_3$plot <- survplot_3$plot +
45 geom_text(data = df, aes(x,y,label = label), color = c("red3",
46 "green2", "blue2"), size = 5)
47
48 survplot_3$plot$labels$y <- "Proportion still on drug" # to change the
49 label
50 survplot_3
51
52
53 Saving surplot
54
55 png("./3_clean_output/figures/PLOT BARI vs TNFi vs OMA curves non
56 adjusted COLOR.png", width = 800, height = 600, res = 100) # opening
57
58
59
60
```

```
1
2
3     graphic device
4
5     survplot_3
6
7     dev.off() # closing graphic device
8
9     Adjusted curves
10
11    # Creating list object
12
13    plots <- list()
14    plots[[1]] <- survplot_1_adj
15    plots[[2]] <- survplot_2_adj
16
17    # Nice function
18
19
20    multi_plot_cox <- arrange_ggsurvplots(plots, print = T, ncol = 2)
21
22    png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA curves
23    adjusted COLOR.png", width = 1000, height = 600, res = 100) # opening
24    graphic device
25
26    multi_plot_cox
27
28
29    dev.off() # closing graphic device
30
31    All curves
32
33    # Creating list object
34
35    plots <- list()
36    plots[[1]] <- survplot_1
37    plots[[3]] <- survplot_2
38    plots[[2]] <- survplot_1_adj
39    plots[[4]] <- survplot_2_adj
40
41    # Nice function
42
43
44    multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2, nrow =
45    2)
46
47    # but does not display properly now.. :(
48
49    AIPTW absolute risk of treatment discontinuation biplot
50
51    plot.ate.diff + plot.ate.diff2
52
53    png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA AIPTW curves
54    adjusted COLOR.png", width = 1000, height = 600, res = 100) # opening
55    graphic device
56
57
58
59
```

```
1
2
3 plot.ate.diff + plot.ate.diff2
4
5 dev.off() # closing graphic device
6
```

7 Diagnostic multipanel plots

8 Asked by Lilly statistician to show balance in this analysis.

```
9
10
11 pscore_plot <- ggplot(test.data, aes(x = pscores, color = cohort, fill
12 = cohort)) +
13   geom_density(alpha = .47) +
14
15   theme_minimal()+
16   theme(axis.ticks.y = element_blank(),
17         panel.grid.minor = element_blank(),
18         legend.title = element_blank(),
19         text = element_text(size = 16),
20         axis.title.x = element_text(hjust = 0.2, size = 16))+
21
22   scale_colour_manual(values = c("red2","green3"))+
23   scale_fill_manual(values = c("red2","green3"))+
24
25   xlab("Probability of being assigned BARI or TNFi") +
26   ylab("Density") +
27   labs(title = "A1 - BARI vs TNFi")
28
29
```

```
30 pscore_plot # overlap
```

```
31
32 pscore_plot2 <- ggplot(test.data2, aes(x = pscores, color = cohort,
33 fill = cohort)) +
34   geom_density(alpha = .47) +
35
36   theme_minimal()+
37   theme(axis.ticks.y = element_blank(),
38         panel.grid.minor = element_blank(),
39         legend.title = element_blank(),
40         text = element_text(size = 16),
41         axis.title.x = element_text(hjust = 0.2, size = 16))+
42
43   scale_colour_manual(values = c("red2","blue3"))+
44   scale_fill_manual(values = c("red2","blue3"))+
45
46   xlab("Probability of being assigned BARI or OMA") +
47   ylab("Density") +
48   labs(title = "A2 - BARI vs OMA")
49
```

```
50
51
52 pscore_plot2
53 # Good overlap
54
55
56
57
58
59
```



```

1
2
3   library(cobalt)
4
5   # BARI vs TNFi
6   B1 <- love.plot(COVS, treat = test.data$cohort, weights =
7   test.data$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
8   var.order = "adjusted", title = "B1 - BARI vs TNFi", color =
9   c("#FD8D3C", "#08306B"), themes = theme_pubclean() )
10
11
12  # BARI vs OMA
13  B2 <- love.plot(COVS_2, treat = test.data2$cohort, weights =
14  test.data2$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
15  var.order = "adjusted", title = "B2 - BARI vs OMA", color =
16  c("#FD8D3C", "#08306B") , themes = theme_pubclean() )
17
18  one <- ( pscore_plot + B1)
19
20  two <- ( pscore_plot2 + B2 )
21
22  png("./3_clean_output/figures/AIPTW diagnostc COLOR.png", width =
23  1300, height = 900, res = 100) # opening graphic device
24
25  one / two
26
27
28  dev.off() # closing graphic device
29
30

```

1. [3] Fist line analysis

Non-adjusted Survival curves

BARI vs TNFi

```

35
36
37  BARI_first1 <- copy(BARI_first[cohort %in% c("BARI", "TNFi")])
38
39
40  surv_object3 <- Surv(time = BARI_first1$time_on_drug, event =
41  BARI_first1$stop_DMARD) # indiqate stop variable and time_on_drug
42  summary(coxph(surv_object3 ~ cohort, data = BARI_first1))
43  fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first1) # function
44  which creates Kaplan-meier data
45  survplot_first1 <- ggsurvplot(fit3, data = BARI_first1, # plot
46  pval = T,
47  pval.method = TRUE,
48  legend.title = "Groups :",
49  legend.labs = c("BARI", "TFNi"),
50  xlab = "Time (days)",
51  xlim = c(0, 700),
52  censor = FALSE,
53  title = "A - BARI vs TNFi",
54  surv.median.line = "v",
55  linetype = 1,
56
57
58
59

```

```
1
2
3     size = 1.5,
4     ggtheme = theme_benoit(),
5     # palette = c("grey78", "grey50"),
6     palette = c("red2", "green3"), # to get colors
7     risk.table = T
8   )
9
10  survplot_first1$plot$labels$y <- "Proportion still on drug" # to
11  change the label
12  survplot_first1
13
14
15  rm(surv_object3, fit3)
16
17  saving plot curves
18
19  png("./3_clean_output/PLOT BARI vs TNFi first line curves non adjusted
20  COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
21  device
22
23  survplot_first1
24
25  dev.off() # closing graphic device
26
27  BARI vs OMA
28
29  BARI_first2 <- BARI_first[line_of_therapy == "1st" & cohort %in%
30  c("BARI", "OMA")] # selection des TC TNFi
31
32
33  surv_object3 <- Surv(time = BARI_first2$time_on_drug, event =
34  BARI_first2$stop_DMARD) # indiquer stop variable and time_on_drug
35  summary(coxph(surv_object3 ~ cohort, data = BARI_first2))
36  fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first2) # fonction
37  which creates Kaplan-meier data
38  survplot_first2 <- ggsurvplot(fit3, data = BARI_first2, # plot
39    pval = T,
40    pval.method = TRUE,
41    legend.title = "Groups :",
42    legend.labs = c("BARI", "OMA"),
43    xlab = "Time (days)",
44    xlim = c(0, 700),
45    censor = FALSE,
46    title = "B - BARI vs OMA",
47    surv.median.line = "v",
48    linetype = 1,
49    size = 1.5,
50    ggtheme = theme_benoit(),
51    # palette = c("grey78", "grey50", "grey10"),
52    palette = c("red2", "blue3"), # to get colors
53    risk.table = T
54  )
55
56
57
58
59
60
```

```

1
2
3
4   survplot_first2$plot$labels$y <- "Proportion still on drug" # to
5   change the label
6   survplot_first2
7
8
9   rm(surv_object3, fit3)
10
11  saving plot curves
12
13  png("./3_clean_output/PLOT BARI vs OMA first line curves non adjusted
14  COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
15  device
16
17  survplot_first2
18
19  dev.off() # closing graphic device
20
21  Adjusted curves with imputed data (BARI vs TNFi)
22  dummy_cox_impute_first1 <- mice::complete(imputed_data1_first, "long",
23  include = T)
24  dummy_cox_impute_first1 <-
25  dummy_cox_impute_first1[dummy_cox_impute_first1$.imp != 0,]
26
27  BARI_first1_fit <- survfit(coxph(Surv(time = time_on_drug, event =
28  stop_DMARD) ~ cohort+
29
30      I(age_base/10)+
31      bmi_base+
32      concomitant_csDMARD+
33      PREDNISON_STEROID+
34      CDAI0+
35      I(disease_duration_base_years/10)+
36      C(smoker_base, base=3)+
37      line_of_therapy+
38      gender+
39      seropositivity_base+
40      cluster(patient_id)+
41      strata(cohort), dummy_cox_impute_first1),
42  data = dummy_cox_impute_first1)
43
44
45
46  survplot_first1_adj <- ggsurvplot(BARI_first1_fit, data =
47  dummy_cox_impute_first1, variable = "cohort",
48      xlab = "Time (days)",
49      title = "Multivariable Cox model of drug discontinuation by
50  type of treatment - 1st line vs 1st line",
51      legend.title = "Groups :",
52      legend.labs = c("Baricitinib", "TNFi"),
53      censor = FALSE,
54      xlim = c(0, 700),
55      surv.median.line = "v",
56
57
58
59
60

```

```
1
2
3     linetype = 1,
4     size = 1.5,
5     ggtheme = theme_minimal(),
6     #palette = c("grey78", "grey10")
7     palette = c("red2", "green3"), # to get colors
8   )
9
10  survplot_first1_adj <- survplot_first1_adj +
11    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
12    concomitant glucocorticoid, baseline CDAI, disease duration (decades),
13    smoking status, line of therapy, gender, seropositivity")
14
15
16  survplot_first1_adj
17  table(BARI_first1$cohort)
18  rm(dummy_cox_impute_first1, BARI_first1_fit)
19
20  Saving the survival plot in PNG file
21
22  png("./3_clean_output/PLOT BARI vs TNFi first line curves adjusted
23  COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
24  device
25
26  survplot_first1_adj
27
28  dev.off() # closing graphic device
29
30  Adjusted curves with imputed data (BARI vs OMA)
31  dummy_cox_impute_first2 <- mice::complete(imputed_data2_first, "long",
32  include = T)
33  dummy_cox_impute_first2 <-
34  dummy_cox_impute_first2[dummy_cox_impute_first2$.imp != 0,]
35
36
37  BARI_first2_fit <- survfit(coxph(Surv(time = time_on_drug, event =
38  stop_DMARD) ~ cohort+
39    I(age_base/10)+
40    bmi_base+
41    concomitant_csDMARD+
42    PREDNISON_STEROID+
43    CDAI0+
44    I(disease_duration_base_years/10)+
45    C(smoker_base, base=3)+
46    line_of_therapy+
47    gender+
48    seropositivity_base+
49    cluster(patient_id)+
50    strata(cohort), dummy_cox_impute_first2),
51  data = dummy_cox_impute_first2)
52
53
54
55  survplot_first2_adj <- ggsvplot(BARI_first2_fit, data =
```

```
1
2
3 dummy_cox_impute_first2, variable = "cohort",
4     xlab = "Time (days)",
5     title = "Multivariable Cox model of drug discontinuation by
6 type of treatment - 1st line vs 1st line",
7     legend.title = "Groups :",
8     legend.labs = c("Baricitinib", "OMA bDMARDs"),
9     censor = FALSE,
10    xlim = c(0, 700),
11    surv.median.line = "v",
12    linetype = 1,
13    size = 1.5,
14    ggtheme = theme_minimal(),
15    #palette = c("grey78", "grey50")
16    palette = c("red2", "blue3"), # to get colors
17    )
18
19
```

```
20 survplot_first2_adj <- survplot_first2_adj +
21     labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
22 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
23 smoking status, line of therapy, gender, seropositivity")
24
```

```
25
26 survplot_first2_adj
27 table(BARI_first2$cohort)
28 rm(dummy_cox_impute_first2, BARI_first2_fit)
29
```

30 Saving the survival plot in PNG file

```
31 png("./3_clean_output/PLOT BARI vs OMA first line curves adjusted
32 COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
33 device
34
```

```
35
36 survplot_first2_adj
37
```

```
38 dev.off() # closing graphic device
39
```

40 Multipanel plots

41 Non adjusted curves

```
42
43 plots <- list()
44 plots[[1]] <- survplot_first1
45 plots[[2]] <- survplot_first2
46
47
```

48 # Nice function

```
49
50 multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
51
```

```
52
53 png("./3_clean_output/figures/BILOT BARI vs TNFi vs OMA 1st Line
54 curves non-adjusted COLOR.png", width = 1000, height = 600, res = 100)
55 # opening graphic device
56
57
58
59
```

```
1
2
3
4 multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
5
6 dev.off() # closing graphic device
7
8 All in one BARI vs TNFi vs OMA
9
10 BARI_first[,time_on_drug_year := time_on_drug/365.25]
11
12 surv_object4 <- Surv(time = BARI_first$time_on_drug_year, event =
13 BARI_first$stop_DMARD)
14 fit4 <- survfit(surv_object4 ~ cohort, data = BARI_first) # this
15 function creates the data for Kaplan Meyer
16
17
18 survplot_4 <- ggsurvplot(fit4, data = BARI_first, # plot
19 pval = F,
20 pval.method = F,
21 legend.title = "Groups",
22 legend.labs = c("BARI", "TNFi", "OMA"),
23 xlab = "Time (years)",
24 xlim = c(0, 2.5),
25 censor = FALSE,
26 # title = "Non-adjusted drug discontinuation by type of
27 treatment (Kaplan-Meier)",
28 surv.median.line = "v",
29 linetype = 1,
30 size = 1.5,
31 ggtheme = theme_benoit(),
32 palette = c("red3", "green2", "blue2"), # to put colors
33 risk.table = T)
34
35
36 values <- summary(fit4)$table[, "median"]
37 df <- data.frame(y = .2, x = values+.2, label =
38 as.character(paste(round(values*365.25, 2), "\n days")))
39 df <- df[2,]
40
41 survplot_4$plot <- survplot_4$plot +
42   geom_text(data = df, aes(x, y, label = label), color = c("green2"),
43   size = 5)
44
45
46 survplot_4$plot$labels$y <- "Proportion still on drug" # to change the
47 label
48 survplot_4
49
50 Saving survplot
51
52 png("./3_clean_output/figures/PLOT BARI vs TNFi vs OMA first curves
53 non adjusted COLOR.png", width = 800, height = 600, res = 100) #
54 opening graphic device
55
56
57
58
59
60
```

```

1
2
3     survplot_4
4
5     dev.off() # closing graphic device
6
7

```

1. [4] LACK of EFFICACY and ADVERSE EVENTS

Analysis by stop_reasons in competing risk

(BARI vs TNFi)

Cumulative incidence function

```

16     ci_long$time_months <- ci_long$time/365.25*12
17
18     plot2 <- ggplot(data = ci_long, aes(x = time_months,
19                                         y = value,
20                                         linetype = variable ,
21                                         col = cohort )) +
22
23         geom_line(size = 0.75)+
24
25         scale_color_manual(breaks = c(0,1),
26                             values = c("red3","green2"),
27                             labels = c("BARI","TNFi"))+
28         scale_linetype_manual(breaks = c("CI.1","CI.2"),
29                               values = c("solid","dashed", "dotted"),
30                               labels = c("Adverse Event","Ineffectiveness"))
31 + # not showing the "other category"
32     scale_x_continuous(name = "Time (months)",
33                       breaks = c(0,3,6,9,12),
34                       limits = c(0,12)) +
35     scale_y_continuous(name = "Cumulative incidence", limits = c(0,0.4))
36 +
37     theme_benoit()+
38     theme(legend.box = "horizontal",
39           legend.position = c(0.05,1),
40           legend.justification = c(0,1),
41           legend.background = element_blank(),
42           legend.key = element_blank(), #no legend key background
43           legend.key.width = grid::unit(2, "lines"))+ # longer line in
44     legend, to see properly the dashed
45     labs(color = "", linetype = "", title = "A - BARI vs TNFi")
46
47
48     plot2
49
50     ggsave(filename = "PLOT BARI vs TNFi cumulative incidence.png",plot =
51     plot2, path = "./3_clean_output/", device = "png", width = 829, height
52     = 550, units = "px", scale = 3.2)
53
54
55
56
57
58
59
60

```

(BARI vs OMA)*Cumulative incidence function*

```

1
2
3
4
5
6 ci_long_2$time_months <- ci_long_2$time/365.25*12
7
8
9 plot3 <- ggplot(data = ci_long_2, aes(x = time_months,
10                                     y = value,
11                                     linetype = variable ,
12                                     col = cohort )) +
13
14   geom_line(size = 0.75)+
15
16   scale_color_manual(breaks = c(0,1),
17                     values = c("red3","blue2"),
18                     labels = c("BARI","OMA"))+
19   scale_linetype_manual(breaks = c("CI.1","CI.2"), # not showing the
20 "other" category
21                       values = c("solid","dashed", "dotted"),
22                       labels = c("Adverse Event","Ineffectiveness"))
23
24 +
25   scale_x_continuous(name = "Time (months)",
26                     breaks = c(0,3,6,9,12),
27                     limits = c(0,12)) +
28   scale_y_continuous(name = "Cumulative incidence", limits = c(0,0.4))
29
30 +
31   theme_benoit()+
32   theme(legend.box = "horizontal",
33         legend.position = c(0.05,1),
34         legend.justification = c(0,1),
35         legend.background = element_blank(),
36         legend.key = element_blank(), #no legend key background
37         legend.key.width = grid::unit(2, "lines"))+ # longer line in
38 legend, to see properly the dashed
39   labs(color = "", linetype = "", title = "B - BARI vs OMA")
40
41 plot3
42
43 ggsave(filename = "./3_clean_output/figures/PLOT BARI vs OMA
44 cumulative incidence.png", plot3, height = 4, width = 6, units =
45 "in",dpi = 300)

```

Multipanel

```

46
47 plot2_3 <- plot2 + plot3
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

59 ggsave(filename = "./3_clean_output/figures/PLOT BARI vs TNFi and BARI
60 vs OMA cumulative incidence.png", plot2_3, height = 4, width = 8,
61 units = "in",dpi = 300)

```


1. [6] LDA - REM

Exploration

See all available raw CDAI measures : :)

```

BARI_long[, group := "non-BARI"]
BARI_long[drug == "BIOLOGIC_BARICITINIB", group := "BARI"]

plot_data <- copy(BARI_long[!is.na(TC_id) & TC_id %in%
BARI_DATA$TC_id])
plot_data <- merge(plot_data, BARI_DATA[,.(TC_id, cohort)], by =
"TC_id")

CDAI_plot <- ggplot(data = plot_data,
                    aes(x = time, y = CDAI, fill = cohort) )+

  annotate("rect",xmin = 0.875,xmax = 1.125,
          ymin = -1,ymax = 60,alpha = .5,fill = "grey80")+
  annotate("rect",xmin = -.05,xmax = 0+1.5/12,
          ymin = -1,ymax = 60,alpha = .5,fill = "grey80")+

  geom_point(data = plot_data[cohort != "BARI"], alpha = 0.2, size =
2, shape = 21, position = position_jitter(width = 0.02, seed = 123) )+
  geom_point(data = plot_data[cohort == "BARI"], alpha = 0.25, size =
2, shape = 21, position = position_jitter(width = 0.02, seed = 123) )+

  #geom_jitter(width = 0.01, height = 0.01, data = plot_data[cohort !=
"BARI"], alpha = 0.2, size = 2, shape = 21, show.legend = F )+
  #geom_jitter(width = 0.01, height = 0.01, data = plot_data[cohort ==
"BARI"], alpha = 0.25, size = 2, shape = 21 , show.legend = F, )+

  geom_smooth(alpha = 0.1, size = 1, aes(color = cohort), show.legend
= F)+

  coord_cartesian(xlim = c(0,2.5))+
  labs(title = "CDAI across time type of treatment (all TC)",
        x = "Time (years since TC initiation)",
        y = "CDAI score",
        color = "",
        fill = "")+
  theme_benoit()+
  theme(legend.position = c(1,1),
        legend.justification = c(1,1))+
  guides(color = guide_legend(override.aes = list(linetype = NA,size =
3)))

CDAI_plot

```

To me this figure is the best results to be discussed regarding REM and LDA

saving plot

```
png("./3_clean_output/figures/PLOT CDAI across time raw.png",  
     width = 8,  
     height = 6,  
     units = "in",  
     res = 120) # opening graphic device
```

CDAI_plot

```
dev.off() # closing graphic device
```

CARRAC histogram

Building large format data table from the CARRAC output

```
# Extracting LDA BARI  
LDA_BARI <- rbind(LDA_BARI_TNF[2,1:4], LDA_BARI_OMA[2,1:4]) # I have  
one estimation per comparison  
LDA_BARI[, LDA := mean(LDA)][, LDA_sup := mean(LDA_sup)][, LDA_inf :=  
mean(LDA_inf)] # averaging  
LDA_BARI <- LDA_BARI[1]  
  
# REM BARI  
REM_BARI <- rbind(REM_BARI_TNF[2,1:4], REM_BARI_OMA[2,1:4]) # I have  
one estimation per comparison  
REM_BARI[, LDA := mean(LDA)][, LDA_sup := mean(LDA_sup)][, LDA_inf :=  
mean(LDA_inf)] # averaging  
REM_BARI <- REM_BARI[1]  
  
# IDem for TNFi and OMA  
LDA_TNFi <- LDA_BARI_TNF[3,1:4]  
REM_TNFi <- REM_BARI_TNF[3,1:4]  
  
LDA_OMA <- LDA_BARI_OMA[3,1:4]  
REM_OMA <- REM_BARI_OMA[3,1:4]  
  
# Binding together  
LDA <- rbind(LDA_BARI, LDA_TNFi, LDA_OMA)  
setnames(LDA, c("ttt", "LDA", "LDA_sup", "LDA_inf")) #putting right  
labels  
REM <- rbind(REM_BARI, REM_TNFi, REM_OMA)  
setnames(REM, c("ttt", "REM", "REM_sup", "REM_inf"))  
  
histo_carrac <- cbind(LDA, REM[,-1])  
  
plotting
```

```

1
2
3 carrac_plot <- ggplot(data = histo_carrac, aes(x = ttt, group = ttt))
4 +
5
6   theme_pubclean()+
7
8   geom_errorbar( mapping=aes(x=ttt, ymin=LDA_inf*100,
9 ymax=LDA_sup*100), width=0.2, size=1, color="grey70")+
10   geom_errorbar( mapping=aes(x=ttt, ymin=REM_inf*100,
11 ymax=REM_sup*100), width=0.2, size=1, color="grey50")+
12
13   geom_bar(aes(y = LDA*100), stat = "identity", fill = "grey80", alpha
14 = 0.5)+
15   geom_text(aes(y = LDA*100, label = "LDA"), vjust = 1.5) +
16
17   geom_bar(aes(x = ttt, y = REM*100), stat = "identity", fill =
18 "grey65", alpha = 0.5)+
19   geom_text(aes(x= ttt, y = REM*100, label = "REM"), vjust=1.5) +
20
21   theme(strip.text.y = element_blank(),
22         strip.background = element_blank(),
23         axis.line.x = element_line(size = 0.5),
24         axis.text = element_text(face = "bold", colour = "black"),
25         legend.position="bottom", plot.margin =
26         unit(c(1,3,2,1),"lines"))+
27
28   scale_y_continuous(limits = c(0,82))+
29
30   labs(y = "(% of TC)", x = "Treatment group", title = "A - REM and
31 LDA rates \nby type of treatment \n(CARRAC)")
32
33 carrac_plot
34
35 also Saving CARRAC plot only
36
37 png("./3_clean_output/figures/PLOT BARI 3 CARRAC ONLY.png", width =
38 350, height = 600, res = 100) # opening graphic device
39
40 ggplot(data = histo_carrac, aes(x = ttt, group = ttt)) +
41
42   theme_pubclean()+
43
44   geom_errorbar( mapping=aes(x=ttt, ymin=LDA_inf*100,
45 ymax=LDA_sup*100), width=0.2, size=1, color="grey70")+
46   geom_errorbar( mapping=aes(x=ttt, ymin=REM_inf*100,
47 ymax=REM_sup*100), width=0.2, size=1, color="grey50")+
48
49   geom_bar(aes(y = LDA*100), stat = "identity", fill = "grey80", alpha
50 = 0.5)+
51   geom_text(aes(y = LDA*100, label = "LDA"), vjust = 1.5) +
52
53
54
55
56
57
58
59

```

```
1
2
3   geom_bar(aes(x = ttt, y = REM*100), stat = "identity", fill =
4 "grey65", alpha = 0.5)+
5   geom_text(aes(x= ttt, y = REM*100, label = "REM"), vjust=1.5) +
6
7   theme(strip.text.y = element_blank(),
8         strip.background = element_blank(),
9         axis.line.x = element_line(size = 0.5),
10        axis.text = element_text(face = "bold", colour = "black"),
11        legend.position="bottom", plot.margin =
12 unit(c(1,3,2,1),"lines"))+
13
14   scale_y_continuous(limits = c(0,82))+
15
16   labs(y = "(% of TC)", x = "Treatment group", title = "REM and LDA
17 rates \nby type of treatment \n(CARRAC)")
18
19
20 dev.off() # closing graphic device
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	p2 p2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	p4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	p4-5
Methods			
Study design	4	Present key elements of study design early in the paper	p6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	p6-7
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	p6 p6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	p7 & supp p2
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	supp p2
Bias	9	Describe any efforts to address potential sources of bias	p16
Study size	10	Explain how the study size was arrived at	supp p9
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	p7-8
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	p8-9 supp p 5-6 p8-9 p8-9 p9 & supp p5-6

Continued on next page

Results

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	p10
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	p11-12
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	p13
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	p13-14
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	p14-15 & supp

Discussion

Key results	18	Summarise key results with reference to study objectives	p15
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	p16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	p17
Generalisability	21	Discuss the generalisability (external validity) of the study results	p17

Other information

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	p18
---------	----	---------------------------------------------------------------------------------------------------------------------------------------------------------------	-----

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

Filled in by Benoît GILBERT, 30-01-2023

BMJ Open

REAL WORLD EFFECTIVENESS OF BARICITINIB IN THE SWISS RHEUMATOID ARTHRITIS REGISTER (SCQM-RA)

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2023-072300.R1
Article Type:	Original research
Date Submitted by the Author:	20-Dec-2023
Complete List of Authors:	Gilbert, Benoit Thomas P.; Geneva University Hospitals, Division of Rheumatology, Department of Internal Medicine Specialties Mongin, Denis; Hopitaux Universitaires de Geneve, Rheumatology, Department of Internal Medicine Specialties Aymon, Romain; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties Lauper, Kim; Hopitaux Universitaires de Geneve, Rheumatology, Department of Internal Medicine Specialties LAEDERMANN, Cédric; Eli Lilly PERRIER, Clémentine; Eli Lilly Mueller, Ruediger; Rheumazentrum Ostschweiz, private practice Courvoisier, Delphine; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties Finckh, Axel; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Epidemiology
Keywords:	RHEUMATOLOGY, EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

REAL WORLD EFFECTIVENESS OF BARICITINIB IN THE SWISS RHEUMATOID ARTHRITIS REGISTER (SCQM-RA)

Benoit GILBERT^{1*}, Denis MONGIN¹, Romain AYMONT¹, Kim LAUPER^{1,2}, Cédric LAEDERMANN³,
Clémentine PERRIER³, Ruediger MUELLER⁴, Delphine S. COURVOISIER¹, Axel FINCKH¹

¹ Division of Rheumatology, Geneva University Hospitals, Geneva.

² Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research,
University of Manchester.

³ Eli Lilly (Suisse) SA, 16 Ch Des Coquelicots, CH-1214 Vernier.

⁴ Private practice, St Gallen, Switzerland.

* Correspondence:

Benoît Thomas P. GILBERT

26 Avenue de Beau-Séjour

Division of Rheumatology, Department of Medicine, Geneva University Hospital,

1206 Geneva,

Switzerland

benoit.gilbert@unige.ch

ABSTRACT

Objectives: This observational study compares the effectiveness of baricitinib (BARI), a targeted synthetic DMARD (tsDMARD) with alternative biological DMARDs (bDMARDs) in rheumatoid arthritis (RA) patients, from a prospective, longitudinal cohort.

Methods: We compared patients initiating a treatment course of BARI, tumor necrosis factor inhibitors (TNFi) or other mode of action bDMARDs (OMA), during a period when all these DMARDs were available in Switzerland. The primary outcome was drug-maintenance; secondary outcomes included discontinuation rates related specifically to ineffectiveness and to adverse events. We further analyzed rates of low disease activity (LDA) and remission (REM) at 12 months, and drug maintenance in b- and tsDMARD-naïve population.

Results: A total of 1053 treatment courses (TC) were included: 273 on BARI, 473 on TNFi and 307 on OMA. BARI was prescribed to older patients with longer disease duration and more previous treatment failures than TNFi. Compared to BARI, the adjusted drug maintenance was significantly shorter for TNFi (hazard ratio (HR) for discontinuation: 1.76; 95% CI [1.32-2.35]), but not compared to OMA (HR 1.27; 95% CI [0.93-1.72]). These results were similar in the b/tsDMARD-naïve population. The higher discontinuation of TNFi was mostly due to an increased discontinuation for ineffectiveness (HR = 1.49; 95% CI [1.03 – 2.15]), with no significant differences in drug discontinuation for adverse events (HR = 1.46; 95% CI [0.83 - 2.57]). The LDA and REM rates at 12 months did not differ significantly between the 3 groups.

Conclusions: BARI demonstrated a significantly higher drug maintenance compared to TNFi, mainly due to lower drug discontinuations for ineffectiveness. We found no difference in drug-maintenance between BARI and OMA. Clinical outcomes did not differ between the three groups. Our results suggest that BARI is an appropriate therapeutic alternative to bDMARDs in the management of RA.

Strengths and limitations of this study

Strengths:

- Use data derived from office-based rheumatologists
- Study period where all alternative medications were available on the market
- Several sensitivity analyses, congruent with main results

Limitations:

- Not a randomized setting
- Sub-analysis in b/tsDMARD-naïve population has limited sample size

INTRODUCTION

Rheumatoid arthritis (RA) is an auto-immune disease leading to widespread inflammation and irreversible joint damage, if insufficiently treated. New treatment paradigms have emerged in the last decades, such as “early aggressive therapy” in the so called “window of opportunity”, during which patients are more likely to reach long term remission.[1] A wide panel of biological disease modifying antirheumatic drugs (bDMARDs) and targeted synthetic DMARDs (tsDMARDs) have been approved in the management of RA, after failure of methotrexate. In clinical-trial settings, b- and tsDMARDs have demonstrated significant reduction of joint inflammation and prevention of joint damage.[2–8]

Efficacy estimates from placebo-controlled randomized trials often differ from real-world effectiveness estimates, because of patient selection, adherence to therapy and other reasons.[9–12] Indeed, drug maintenance of many bDMARDs remains modest in observational analyses, while long term remissions are rare and secondary loss of efficacy frequent.[13] Furthermore, understanding the clinical effectiveness of bDMARDs or tsDMARDs in specific conditions, such as elderly or multi-morbid patients, may become important as we move towards personalized care. Finally, trials provide only limited data on long term effectiveness and safety because clinical-trial follow-up is typically less than 12 months.

Baricitinib (BARI) has been approved in Switzerland for the treatment of RA in 2017 as well as all around the world. Clinical trials with BARI have established efficacy and demonstrated acceptable adverse events profile, both in combination with methotrexate or in monotherapy.[14–20] However, evidence about effectiveness of BARI compared to TNFi in

1
2
3 real-world settings are scarce. A recently published analysis of registry data from Sweden
4
5 showed that baricitinib had higher maintenance as compared to most other bDMARD.[21]
6
7 Pappas et al., in the United States, also demonstrated that TNFi and non-TNFi drugs had
8
9 similar outcomes when prescribed in b/tsDMARD-naïve population, an observation replicated
10
11 in the RA-BE-REAL study.[22,23]
12
13

14
15 The aim of our analysis was to compare real-world drug maintenance between BARI and other
16
17 approved b/tsDMARDs, using data from a European registry.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODS

Study population

This is a nested cohort study from a prospective, longitudinal, cohort of Swiss RA patients in a real-life setting, the Swiss Clinical Quality Management registry (SCQM). The SCQM registry was founded in 1997 with the financial support of Swiss regulatory authorities, who recommended a continuous monitoring of all patients receiving new DMARDs. Unlike many other European registries, most patients are enrolled by private office-based rheumatologists (60%), providing a population-based sample of RA patients in Switzerland. All approved RA treatments are represented in the registry. The data for this analysis was extracted from the SCQM registry on 2020-06-01.

We used “treatment courses” (TCs) as our denominator of interest, with each new treatment initiation considered as a separate “treatment course” (TC). We included all TCs with the medications of interest initiated between 2017-09-01 and 2020-06-01, with at least one follow-up visit, in adult patients with a diagnosis of RA confirmed by a rheumatologist. Thus, a given patient could potentially contribute to several TCs during the study period. To minimize the risk of confounding bias, the time window was selected to include only the period when all the therapies examined were available for prescription and reimbursed (BARI was first reimbursed on the Swiss market in September 2017). We excluded TCs with no follow-up visit at the time of data extraction.

Exposure of interest

The exposure of interest was the type of treatment used, namely BARI, TNFi, and other mode of action bDMARDs (OMA), excluding other tsDMARDs and rituximab. We decided to exclude rituximab a priori, because its long-term action impairs precise estimation of treatment discontinuation. Tofacitinib was excluded because we had insufficient TC to perform meaningful comparative effectiveness analyses against a single other specific tsDMARD agent. Included TNFi treatments were: adalimumab, etanercept, golimumab, certolizumab, infliximab. Included OMA treatments were: tocilizumab, abatacept, sarilumab, and anakinra.

Outcomes

The primary outcome of this analysis was the time to all-cause-discontinuation. This outcome, also referred to as “drug maintenance”, captures both the drug’s effectiveness and its tolerance.[24] The time to all-cause-discontinuation was defined as the number of days between treatment initiation and the reported date of discontinuation, or the date of initiation of a new b/tsDMARD, whatever came first. In survival analyses, death or lost-to-follow up are censored. We also report discontinuation rates at 12 months. Temporary discontinuations of less than 6 months (for instance, because of an elective surgery or a pregnancy) were not considered a permanent drug discontinuation. Discontinuation reasons are recorded by the clinician when stopping a DMARD treatment, who chooses between four options (“Adverse event”, “Ineffectiveness”, “Remission”, or “Other”).

Pre-planned secondary outcomes, were time to discontinuation due to ineffectiveness and time to discontinuation due to adverse events. Other secondary outcomes included response rates, namely the rates of low disease activity (LDA) and remission (REM), at 12 months, defined respectively as attaining a CDAI score ≤ 10 and CDAI score ≤ 2.8 (not mutually

1
2
3 exclusive).[25] Finally, we performed an exploratory subgroup analysis, restricting the
4
5 population to b/tsDMARD-naive patients only, and re-assessing the main outcome in this
6
7 setting.
8
9

10 **Statistical analysis**

11
12
13 Analyses were conducted and reported in accordance to EULAR recommendations for
14
15 comparative effectiveness research.[9] Baseline characteristics were compared using
16
17 generalized linear mixed models to account for repeated treatments within the same patients.
18
19 For the primary outcomes, Kaplan-Meier survival analyses were used to assess crude drug
20
21 maintenance, and groups were compared using Log-Rank tests. Subsequently, missing
22
23 covariates were imputed using chained equations (see below for details). We then
24
25 implemented Cox proportional hazard ratio models, to obtain adjusted estimates. Based on
26
27 prior subject matter knowledge,[26] we adjusted our models for the following potential
28
29 confounders: age, gender, BMI, concomitant csDMARDs use (yes/no), concomitant
30
31 prednisone usage (yes/no), CDAI score at baseline, disease duration, smoking status (current-,
32
33 former-, never-smoker) and line of therapy (1st, 2nd, 3rd, 4th or more), and seropositivity
34
35 (yes/no). Detailed definitions for each variable are available in the supplement
36
37 (Supplementary Material 1). The main analysis (survival analysis) accounted for clustering
38
39 resulting from patients with multiple treatment courses, inducing correlation within the
40
41 patient-level data. The cluster term is used to compute a robust variance for the model, by
42
43 applying the so-called Huber sandwich estimator.[27] . All conditions of application of the Cox
44
45 model were verified. One additional sensitivity analysis was conducted for the primary
46
47 outcome, using augmented inverse probability of treatment weighting (AIPTW).
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 In secondary analyses, we used the Fine-Gray approach to assess specific reasons for drug
4 discontinuation (i.e. ineffectiveness, or adverse event) in a competing-risk setting. The Fine-
5 Gray method takes competing risks into account when estimating the cumulative incidence
6 function, modelling the sub-distribution hazard without treating competing events as
7 censoring events.[28] Other secondary outcomes included response rates (LDA and REM) at
8 12 months. To avoid overestimations, we computed the response rates using the ‘confounder-
9 adjusted response rate with attrition correction’ (CARRAC) method.[29] The latter estimates
10 the response rates using multiple imputations, with a model including both confounders and
11 treatment stop reason. CARRAC thus provides reliable estimates when reasons for treatment
12 discontinuation differ between compared groups.

13
14 For all adjusted analyses, missing baseline covariates were imputed using closest value in a
15 window of -90 to +30 days. However, this window was reduced to -30 to +7 days when
16 imputing baseline CDAI. If still missing after this first step, baseline CDAI values were imputed
17 using linear mixed effect regression model with quadratic time. We imputed other baseline
18 covariates with chained equations technique, which provides unbiased estimates if the
19 variables are missing at random.[29] Such imputations were performed using 50 datasets with
20 25 iterations. Imputation was done using the whole data set, before adequately subsetting
21 the data for each group comparison.

22
23 We also imputed data required for secondary outcomes, including disease activity. If the CDAI
24 score at 12-month was not available, the closest value in a window of +/- 45 days was used (a
25 3-month-wide window). If still missing, the 12-month CDAI values were imputed using nearest
26 neighboring value, as previously described.[30]

27
28 All analyses were conducted using R (version 4.0.3), in particular with packages “*tableone*”,
29 “*survival*”, “*mice*”.[31] Two tailed p-values < 0.05 were considered significant. We did not
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 adjust p value for multiple comparisons, as outcomes were pre-specified. Final analysis code
4
5 is shown in the supplement (Supplementary Material 2).
6
7

8 **Patient and Public Involvement**

9

10 Patient involvement is central to the SCQM cohort. Several patients are part of the executive
11
12 board and involved in the approval of research projects.
13
14
15
16

17 **RESULTS**

18

19 **Population description**

20
21

22 During the study period, 1053 TC were initiated in 834 different patients, including 273 TC
23
24 with BARI, 473 with TNFi and 307 with OMA (Figure 1; Figure S1). TNFi were more often given
25
26 as a second line therapy after methotrexate failure. Inversely, BARI was prescribed to
27
28 significantly older patients, with longer disease durations and more previous treatment
29
30 failures (Table 1).
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Baseline characteristics of study population, SCQM-RA registry, 2017-2020.

Variable	BARI (TC = 273; 273 pts)		TNFi (TC = 473; 408 pts)		OMA (TC = 307; 289 pts)		p values
	n % of total in group Otherwise: Mean (SD)						
Patients		Miss.		Miss.		Miss.	
Female	78 %	0	74 %	1	73 %	1	0.097
Age (years)	59 (14)	0	52 (15)	1	59 (13)	1	0.021
Disease duration (years)	13 (10)	4	8 (9)	19	11 (9)	5	0.027
CDAI baseline (raw data)	19 (10)	175	18 (10)	301	20 (13)	204	0.34
CDAI baseline (imputed)	15 (9)	0	14 (9)	0	16 (11)	0	0.06
Obesity (BMI > 30)	16 %	104	14 %	134	13 %	115	0.85
Smoking	17 %		18 %		21 %		Ref.
Current	28 %	32	26 %	69	28 %	26	0.95
Former	43 %		41 %		48 %		0.98
Never							0.92
Seropositive (ACPA or RF)	75 %	1	70 %	7	77 %	5	
TC		Miss.		Miss.		Miss.	
Concomitant csDMARD	40 %	0	61 %	0	46 %	0	<0.01
Line of Therapy	17 %		48 %		22 %		Ref.
1 st (= bio-naïve)	20 %	0	23 %	0	24 %	0	<0.01
2 nd	19 %		11 %		24 %		<0.01
3 rd	44 %		18 %		31 %		<0.01
4 th or later							<0.01
Previous tsDMARD use (non-BARI)	33 %	0	1 %	0	5 %	0	<0.01
Concomitant glucocorticoid (at any time)	22 %	0	20 %	0	24 %	0	0.35
Mean dose of concomitant glucocorticoid (mg)	2.0 (4.6)	0	2.1 (5.4)	0	2.2 (5.1)	0	0.95
Dose of BARI (4mg)	86 %	0	-	-	-	-	-

Table 1 Legend: In Switzerland, BARI was prescribed to older patients, with longer disease duration and more previous treatment failures. Missing values for covariables are reported as absolute numbers

BARI = baricitinib. TC = Treatment Courses. CDAI = Clinical Disease Activity Index. TNFi = Tumor Necrosis Factor Inhibitors. OMA = bDMARDs with Other Mode of Action. tsDMARD = targeted synthetic DMARDs. ACPA = Anti Citrullinated Peptide Antibodies. RF = Rheumatoid Factor. Miss. = number of missing values. Ref. = Reference

1
2
3 category for p-values (pairwise comparisons). P-values are obtained by generalized linear mixed models to
4 account for repeated treatments within the same patients. In TFNi and OMA groups, some patients have
5 contributed several TC, thus total number of TCs exceeds total number of patients.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Time to all-cause-discontinuation

Table 2: Crude treatment discontinuation by group and by reason, SCQM, 2017-2020.

	BARI (TCs = 273; 273 pts)	TNFi (TCs = 473; 408 pts)	OMA (TCs = 307; 289 pts)
Treatment discontinuation (all causes)	30 %	43 %	35 %
For adverse events	8 %	10 %	10 %
For ineffectiveness	16 %	23 %	17 %
For remission	0 %	1 %	0 %
For other reason	5 %	8 %	7 %
<p>Table 2 legend: % are computed on total number of TCs per group, for the whole study period. BARI = baricitinib. TNFi = TNF inhibitors. OMA = Other Mode of Action drugs. TC = Treatment Courses. Pts = Number of patients. Due to rounding, the sum of the percentages of the causes of discontinuation may not correspond exactly to the total treatment discontinuation percentage.</p>			

Crude proportions of treatment discontinuation by reasons are reported in Table 2, and crude times of observation are represented on Figure S1.

At 12 months, based on the Kaplan-Meier curves (Figure 2), the estimated proportions of patients still on therapy were: 71% (95% CI [65% - 77%]) in the BARI group, 55% (95% CI [50% - 61%]) in the TNFi group, and 63% (95% CI [57% - 70%]) in the OMA group.

Overall, unadjusted time to all-cause-discontinuation was significantly longer in the BARI group compared to the TNFi group (estimated median prescription survival-time of 704 versus 448 days; Log-rank $p < 0.01$; Figure 2). These results persisted after adjustment for confounding factors using the multivariable Cox model (HR = 1.76; 95% CI [1.32-2.35]; $p < 0.001$; Table S1 and Figure S2; Figure S3).

BARI versus OMA time to all-cause-discontinuation was not significantly different, even after adjustment (HR 1.27; 95% CI [0.93-1.72]; $p = 0.13$; Table S1, Figure S2 and Figure S3).

1
2
3 Sensitivity analyses using AIPTW led to similar conclusions (Figure S4). Covariates significantly
4 associated with decreased drug maintenance were high baseline CDAI scores and concomitant
5 glucocorticoid usage (Table S1 and Figure S2).
6
7
8
9

10 **Time to all-cause-discontinuation in b/tsDMARD-naïve patients**

11
12
13 In this exploratory subgroup analysis, we restricted the population to patients without prior
14 experience of b/tsDMARDs (so-called 'bio-naïve' patients, i.e. first b/tsDMARD prescription
15 after methotrexate failure). In this subpopulation, patient characteristics were more
16 balanced than in the main analysis, except for age, which remained younger in TNFi
17 population, and concomitant csDMARDs usage (more frequent in TNFi) (Table S2). Of note,
18 the sample size was consequently reduced to 46 BARI, 225 TNFi and 66 OMA.
19
20
21
22
23
24
25
26

27 When analysing only these b/tsDMARD-naïve patients, both the non-adjusted (Figure 3) and
28 the adjusted differences between BARI and TNFi became larger (HR TNFi vs BARI = 2.5; 95% CI
29 [1.23 – 5.16]; p=0.01), but the differences between baricitinib and OMA group remained not
30 significantly different (HR OMA vs BARI = 1.90; 95% CI [0.71 – 5.1]; p=0.2).
31
32
33
34
35
36

37 **Time to discontinuation for adverse events or ineffectiveness**

38
39
40 A secondary outcome was the cumulative incidence of drug discontinuation by specific
41 reasons for discontinuation (ineffectiveness or adverse events, Figure 4). Using Fine-Gray
42 adjusted approach, we found no difference in the incidence of adverse event comparing BARI
43 to TNFi (HR = 1.46; 95% CI [0.83 - 2.57]; p=0.13), or BARI to OMA (HR = 1.34; 95% CI [0.74 -
44 2.42]; p=0.25). The incidence of drug discontinuation for ineffectiveness was more frequent
45 in TNFi compared to BARI (HR = 1.49; 95% CI [1.03 – 2.15]; p=0.01), but similar between OMA
46 and BARI (HR = 1.09; 95% CI [0.72 – 1.64]; p=0.69).
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Remission and low disease activity at 12 months

The estimated 12-month rates of REM and LDA, estimated using CARRAC did not differ significantly between the 3 groups (Figure 5). LDA ranged from 62% to 71% and REM ranged from 17% to 26%.

DISCUSSION

In this study, the overall drug maintenance of BARI was significantly longer compared to TNFi, despite the fact that it was prescribed to older patients, with longer disease duration, and more previous treatment failures similar to what was observed in RA-BE-REAL, another real-world study.[23] However, the adjusted 12-month response rates in terms of LDA and REM did not differ significantly between BARI, TNFi and OMA groups. The difference in drug discontinuation owes mainly to more treatment discontinuations for ineffectiveness in the TNFi group compared to the BARI group, while drug discontinuation due to adverse event did not differ significantly between the groups.

Our results are in line with previous findings comparing other JAK-inhibitors (JAKi) (i.e. tofacitinib as well as BARI) to TNFi and OMA medications,[22,32] which reported a longer drug maintenance of tsDMARD compared to TNFi, and similar maintenance to other bDMARDs. Of note, Lauper et al., using data from 19 national registers, found no difference in retention time between JAK-inhibitors and TNFi.[33] Still, Lauper et al. grouped all JAKi together in their study, thus it is not clear if these observations remain true for BARI alone, which might differ from other JAKi. For instance, Barbulescu et al. reported a higher drug maintenance for BARI as compared to tofacitinib.[21]

It was previously shown that BARI is more efficient in relieving pain as compared to adalimumab therapy [34] and some molecular mechanisms relevant to JAK-STAT signalling

1
2
3 have been hypothesized.[35] This observation has been hypothesised to result anti-
4
5 nociceptive effect independent from inflammation.[35] This faster pain relief could partially
6
7 explain why BARI has increased maintenance than other medication in our study, even though
8
9 having similar 12-months LDA and REM rates. An alternative hypothesis is that the more
10
11 convenient oral administration encourages patients to stay on medication longer. Yet, a third
12
13 possible interpretation is that patients who experienced numerous treatment failures tend to
14
15 stay on their latest therapy; however, our study accounts for this potential bias, by performing
16
17 a sensitivity analysis in a subgroup of b-tsDMARD naïve patients, which showed a similar
18
19 result. Finally, given recent discussion regarding tofacitinib safety,[36] future research needs
20
21 to clarify whether a class effect for JAKi related adverse events exist. In this analysis, we found
22
23 no indication for an increased incidence of adverse-related treatment discontinuation with
24
25 BARI compared to alternative bDMARDs. Randomized controlled trials are ongoing to further
26
27 compare safety profile of BARI versus TNFi (NCT04086745 and NCT03915964).
28
29
30
31
32
33

34 35 **Limitations and Strengths**

36
37 This work has several limitations, mostly inherent to the observational setting. First, as this is
38
39 a non-randomized study, we cannot formally exclude unmeasured confounding between the
40
41 groups. The available baseline variables were, in most cases, adequately balanced, except for
42
43 age. When we restricted the analysis to the subgroup of b/tsDMARD naïve patients, we found
44
45 largely similar results. Despite being limited by the small sample size, this exploratory
46
47 subgroup analysis suggests that confounding by line of treatment was adequately accounted
48
49 for in the adjusted analysis.
50
51
52

53
54 Secondly, the average length of follow-up was only approximately 200 days per TC (Figure
55
56 S1). Indeed, our study covers about 2 and a half years, and we only included TC newly initiated
57
58 during this time-windows. Also, because of the study setting, as much as 65% of TC did not
59
60

1
2
3 have CDAI scores recorded at the date of initiation, and many were missing at the 12-month
4
5 exact timepoint (Figure S5). Hence, our analysis of response rates relied heavily on linear
6
7 interpolation techniques, using other available timepoints, which results in large confidence
8
9 intervals for estimated response rates.[30]
10
11

12
13 The main strength of the study is that it relies on real-world data, and includes a relatively
14
15 large number of patients. As these patients are mostly treated by office-based
16
17 rheumatologists, our study population is representative of routine clinical practice. Also, sub-
18
19 group analyses and sensitivity analyses were consistent with the main results.
20
21

22 23 **CONCLUSIONS**

24
25 In this non-randomized cohort study, drug maintenance of BARI was significantly higher than
26
27 TNFi. However, we found no difference in drug maintenance when comparing BARI with other
28
29 bDMARDs. Based on available data, the estimated 12-month response rates did not
30
31 significantly differ between BARI, TNFi and OMA groups. We found no difference in treatment
32
33 discontinuation for adverse event between the three groups. Overall, our results are in line
34
35 with findings from randomized trials..
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

OTHERS

Contributorship statement

Benoît GILBERT contributed to data-management, data-analysis, figures, manuscript drafting.

Denis MONGIN contributed to data-management, data-analysis, figures, manuscript-revision.

Romain AYMOND contributed to data-analysis (in particular, sensitivity analyses), manuscript revision. Kim LAUPER took part in data-analysis and manuscript revision. Céline LAMACCHIA was involved in the study design and manuscript revision. Clémentine PERRIER was involved in the study design and manuscript revision. Ruediger MUELLER contributed to the study design and manuscript revision. Delphine COURVOISIER was involved in the study design, data-analysis and interpretation, manuscript revision. Axel FINCKH was in charge of the study design (principal investigator), data analysis, data interpretation and manuscript revision.

Competing interests

Benoît GILBERT has been once a paid speaker (Eli Lilly) and participated in advisory board (Janssen). Clementine PERRIER is employed by Eli Lilly and holds stock options (Eli Lilly and Company). Cedric LAEDERMANN is employed by Eli Lilly and holds stock options (Eli Lilly and Novartis). Axel FINCKH has received grants or contracts (Eli-Lilly, Pfizer, Abbvie, Gilead, BMS), consulting fees (Astra-Zeneca, Abbvie, Pfizer, Gilead), honorary payments (BMIS, Abbvie, Eli Lilly, Pfizer, MSD), and participated in advisory boards (Astra-Zeneca, Gilead, Novartis, Abbvie, Eli Lilly, Pfizer, J&J, Mylan, UCB). Denis MONGIN, Romain Aymond, Rüdiger Müller and Delphine COURVOISIER have no conflicts of interest to disclose.

Funding

This analysis has been made possible by financial support of Eli Lilly (Suisse) SA to the Geneva University Hospitals (HUG). Also, the SCQM is financially supported by pharmaceutical

1
2
3 industries and donors, including Eli Lilly. A list of financial supporters can be found on
4
5 www.scqm.ch/sponsors .
6
7

8 **Data sharing statement**

9

10 Restrictions apply to the availability of these data. Data is owned by a third party, the Swiss
11
12 Clinical Quality Management in Rheumatic Diseases (SCQM) foundation. Data may be
13
14 obtained after approval and permission from this license holder (SCQM). Contact information
15
16 for data request: scqm@hin.ch
17
18
19

20 **Ethical Review and Regulatory Considerations**

21

22
23 This observational study has been approved by the Geneva ethical review boards (ERBs) as
24
25 required by local law (Project ID: 2019-00930 ; approval date 28 May 2019). Every participant
26
27 has signed an information and consent form at inclusion in the SCQM registry. Hence, this
28
29 study has been conducted in accordance with the ethical principles of the Declaration of
30
31 Helsinki and is consistent with Good Pharmacoepidemiology Practices (GPPs).
32
33
34

35 **Acknowledgements**

36

37
38 Almut Scherer, from the SCQM. Also, all of the participating physicians of the SCQM in
39
40 Rheumatic Diseases. A complete list of rheumatology offices and hospitals that are
41
42 contributing to the SCQM registries can be found on www.scqm.ch/institutions .
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

- 1 Emery P, Seto Y. Role of biologics in early arthritis. *Clin Exp Rheumatol*. 2003;21:S191-194.
- 2 Lipsky PE, van der Heijde DMFM, St. Clair EW, *et al*. Infliximab and Methotrexate in the Treatment of Rheumatoid Arthritis. *New England Journal of Medicine*. 2000;343:1594–602.
- 3 Breedveld FC, Emery P, Keystone E, *et al*. Infliximab in active early rheumatoid arthritis. *Annals of the Rheumatic Diseases*. 2004;63:149–55.
- 4 Klareskog L, van der Heijde D, de Jager JP, *et al*. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *The Lancet*. 2004;363:675–81.
- 5 Weinblatt ME, Kremer JM, Bankhurst AD, *et al*. A Trial of Etanercept, a Recombinant Tumor Necrosis Factor Receptor:Fc Fusion Protein, in Patients with Rheumatoid Arthritis Receiving Methotrexate. *New England Journal of Medicine*. 1999;340:253–9.
- 6 Maini R, St Clair EW, Breedveld F, *et al*. Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. *The Lancet*. 1999;354:1932–9.
- 7 Breedveld FC, Weisman MH, Kavanaugh AF, *et al*. The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis & Rheumatism*. 2006;54:26–37.
- 8 Moreland LW, Schiff MH, Baumgartner SW, *et al*. Etanercept Therapy in Rheumatoid Arthritis. *Annals of Internal Medicine*. 1999;130:478–86.
- 9 Dixon WG, Carmona L, Finckh A, *et al*. EULAR points to consider when establishing, analysing and reporting safety data of biologics registers in rheumatology. *Annals of the Rheumatic Diseases*. 2010;69:1596–602.
- 10 Wolfe F, Michaud K. The Hawthorne Effect, Sponsored Trials, and the Overestimation of Treatment Effectiveness. *The Journal of Rheumatology*. 2010;37:2216–20.
- 11 Wolfe F, Michaud K. The loss of health status in rheumatoid arthritis and the effect of biologic therapy: a longitudinal observational study. *Arthritis Res Ther*. 2010;12:R35.
- 12 Østergaard M, Unkerskov J, Linde L, *et al*. Low remission rates but long drug survival in rheumatoid arthritis patients treated with infliximab or etanercept: results from the nationwide Danish DANBIO database. *Scandinavian Journal of Rheumatology*. 2007;36:151–4.

- 13 Souto A, Maneiro JR, Gómez-Reino JJ. Rate of discontinuation and drug survival of biologic therapies in rheumatoid arthritis: a systematic review and meta-analysis of drug registries and health care databases. *Rheumatology*. 2016;55:523–34.
- 14 Taylor PC, Keystone EC, van der Heijde D, *et al*. Baricitinib versus Placebo or Adalimumab in Rheumatoid Arthritis. *New England Journal of Medicine*. 2017;376:652–62.
- 15 Fleischmann R, Schiff M, Heijde D van der, *et al*. Baricitinib, Methotrexate, or Combination in Patients With Rheumatoid Arthritis and No or Limited Prior Disease-Modifying Antirheumatic Drug Treatment. *Arthritis & Rheumatology*. 2017;69:506–17.
- 16 Dougados M, van der Heijde D, Chen Y-C, *et al*. Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. *Ann Rheum Dis*. 2017;76:88–95.
- 17 Genovese MC, Kremer J, Zamani O, *et al*. Baricitinib in Patients with Refractory Rheumatoid Arthritis. *N Engl J Med*. 2016;374:1243–52.
- 18 Smolen JS, Genovese MC, Takeuchi T, *et al*. Safety Profile of Baricitinib in Patients with Active Rheumatoid Arthritis with over 2 Years Median Time in Treatment. *J Rheumatol*. 2019;46:7–18.
- 19 Keystone EC, Taylor PC, Tanaka Y, *et al*. Patient-reported outcomes from a phase 3 study of baricitinib versus placebo or adalimumab in rheumatoid arthritis: secondary analyses from the RA-BEAM study. *Ann Rheum Dis*. 2017;76:1853–61.
- 20 Taylor PC, Takeuchi T, Burmester GR, *et al*. Safety of baricitinib for the treatment of rheumatoid arthritis over a median of 4.6 and up to 9.3 years of treatment: final results from long-term extension study and integrated database. *Ann Rheum Dis*. 2022;81:335–43.
- 21 Barbulescu A, Askling J, Chatzidionysiou K, *et al*. Effectiveness of baricitinib and tofacitinib compared with bDMARDs in RA: results from a cohort study using nationwide Swedish register data. *Rheumatology*. 2022;keac068.
- 22 Pappas DA, John GS, Etzel CJ, *et al*. Comparative effectiveness of first-line tumour necrosis factor inhibitor versus non-tumour necrosis factor inhibitor biologics and targeted synthetic agents in patients with rheumatoid arthritis: results from a large US registry study. *Annals of the Rheumatic Diseases*. 2021;80:96–102.
- 23 Alten R, Burmester GR, Cerinic MM, *et al*. Pos0666 a Multinational, Prospective, Observational Study in Patients with Rheumatoid Arthritis Receiving Baricitinib, Targeted Synthetic or Biologic Disease-Modifying Therapies: 12 Month Time to Discontinuation, Effectiveness and Patient Reported Outcome Data from the European Cohort. *Annals of the Rheumatic Diseases*. 2022;81:606–7.
- 24 Pan SMD, Scherer A, Gabay C, *et al*. Differential drug retention between anti-TNF agents and alternative biological agents after inadequate response to an anti-TNF agent in rheumatoid arthritis patients. *Annals of the Rheumatic Diseases*. 2012;71:997–9.

- 1
2
3 25 Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease
4 Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis. ;9.
5
6
7 26 Lauper K, Finckh A. Predictive factors of treatment persistence in rheumatoid arthritis.
8 *Joint Bone Spine*. 2020;87:531–4.
9
10 27 White H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test
11 for Heteroskedasticity. *Econometrica*. 1980;48:817–38.
12
13 28 Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing
14 Risk. *Journal of the American Statistical Association*. 1999;94:496–509.
15
16
17 29 Mongin D, Lauper K, Finckh A, *et al*. Accounting for missing data caused by drug cessation
18 in observational comparative effectiveness research: a simulation study. *Ann Rheum Dis*.
19 2022;annrheumdis-2021-221477.
20
21 30 Mongin D, Lauper K, Turesson C, *et al*. Imputing missing data of function and disease
22 activity in rheumatoid arthritis registers: what is the best technique? *RMD Open*.
23 2019;5:e000994.
24
25
26 31 Buuren S van, Groothuis-Oudshoorn K. Mice : Multivariate Imputation by Chained
27 Equations in R. *J Stat Soft*. 2011;45. doi: 10.18637/jss.v045.i03
28
29
30 32 Finckh A, Tellenbach C, Herzog L, *et al*. Comparative effectiveness of antitumour necrosis
31 factor agents, biologics with an alternative mode of action and tofacitinib in an
32 observational cohort of patients with rheumatoid arthritis in Switzerland. *RMD Open*.
33 2020;6:e001174.
34
35
36 33 Lauper K, Iudici M, Mongin D, *et al*. Effectiveness of TNF-inhibitors, abatacept, IL6-
37 inhibitors and JAK-inhibitors in 31 846 patients with rheumatoid arthritis in 19 registers
38 from the 'JAK-pot' collaboration. *Annals of the Rheumatic Diseases*. 2022;81:1358–66.
39
40
41 34 Taylor PC, Lee YC, Fleischmann R, *et al*. Achieving Pain Control in Rheumatoid Arthritis
42 with Baricitinib or Adalimumab Plus Methotrexate: Results from the RA-BEAM Trial. *JCM*.
43 2019;8:831.
44
45
46 35 Busch-Dienstfertig M, González-Rodríguez S. IL-4, JAK-STAT signaling, and pain. *JAKSTAT*.
47 2013;2:e27638.
48
49
50 36 Álvaro-Gracia JM, García-Llorente JF, Valderrama M, *et al*. Update on the Safety Profile of
51 Tofacitinib in Rheumatoid Arthritis from Clinical Trials to Real-World Studies: A Narrative
52 Review. *Rheumatol Ther*. 2021;8:17–40.
53
54
55
56
57
58
59
60

FIGURE LEGEND

Figure 1 - Selection of eligible treatment courses, SCQM, 2017-2020.

Selection of Treatment courses included in final analysis. TC = Treatment Courses. RA = Rheumatoid Arthritis. bDMARD = biological DMARD. tsDMARD = targeted synthetic DMARD. b/tsDMARD = biological and/or targeted synthetic DMARD. TNFi = TNF inhibitors. OMA = Other Mode of Action bDMARDs.

Figure 2 - Non-adjusted time to drug discontinuation analyses (Kaplan-Meier), SCQM, 2017-2020.

These “survival curves” represent the drug maintenance after initiation, as the estimated proportion of patients still on therapy, by treatment group. Death and loss to follow-up were censored. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs. Log-Rank BARI vs TNFi : $p < 0.001$. Log-Rank BARI vs OMA: $p = 0.11$.

Figure 3 - Unadjusted time to drug discontinuation in b/tsDMARD-naïve patients, SCQM, 2017-2020.

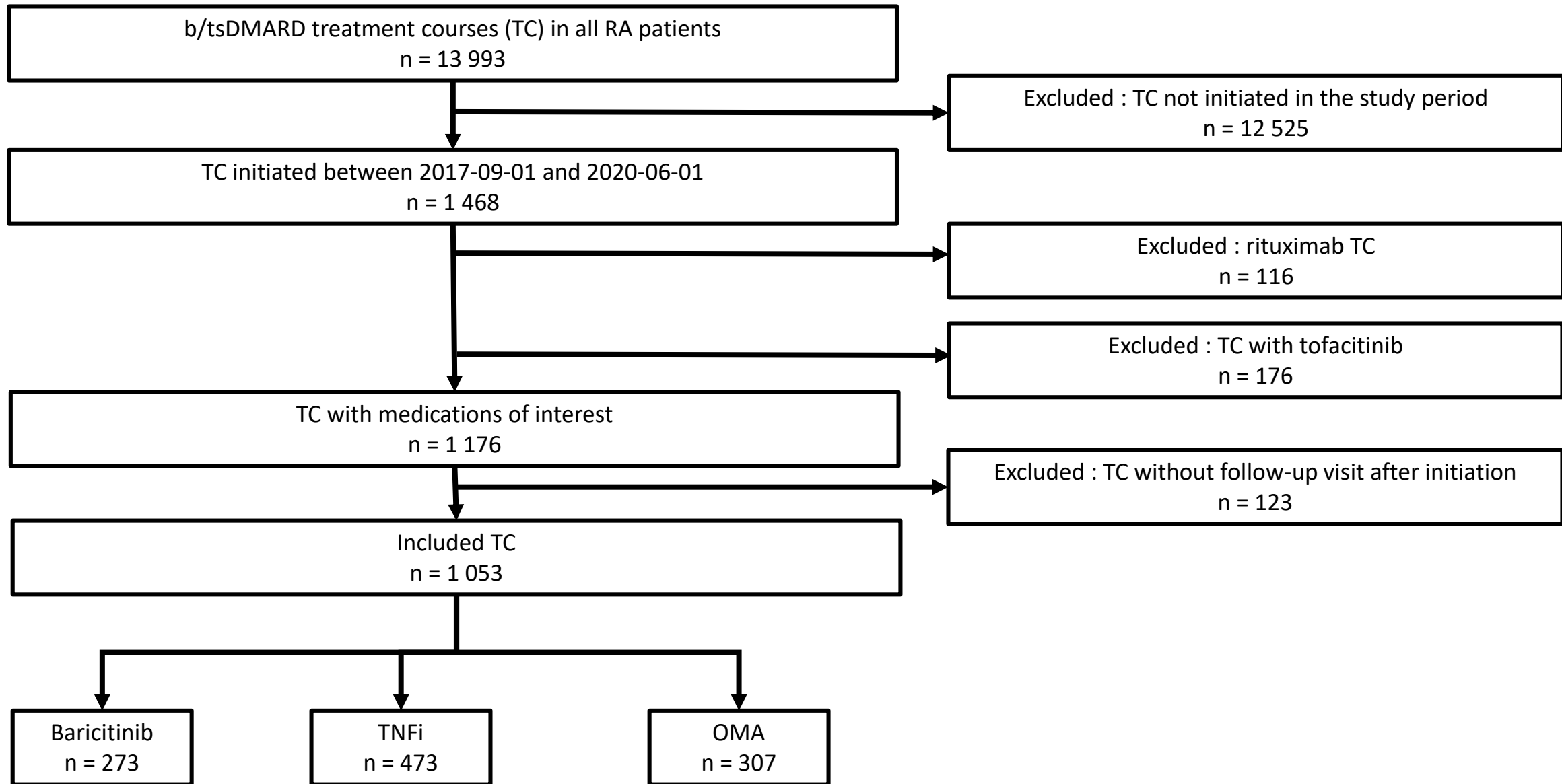
These Kaplan-Meier curves represent the crude “survival” of drug prescription, by treatment group. Death and loss to follow-up are censored. BARI = baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs. Log-Rank BARI vs TNFi : $p = 0.003$. Log-Rank BARI vs OMA : $p = 0.15$.

1
2
3 Figure 4 - Cumulative incidence of drug discontinuation by stop reason and by type of
4 treatment, SCQM, 2017-2020.
5
6

7
8 This figure represents the unadjusted cumulative incidence of drug discontinuation, by group
9 and by reason of discontinuation. BARI = baricitinib. TNFi = Tumor Necrosis Factor. OMA =
10 Other mode of Action bDMARDs.
11
12
13
14
15
16
17
18
19

20 Figure 5 - Estimated response rates at 12-months (CARRAC), SCQM, 2017-2020.
21

22 BARI = baricitinib. TNFi = Tumor Necrosis Factors Inhibitors. OMA = Other Mode of Action
23 bDMARDs. LDA = Low Disease Activity (i.e. CDAI score ≤ 10), in light grey. REM = Remission
24 (i.e. CDAI score ≤ 2.8), in dark grey. 95% CI are represented. This method does not allow
25 computing p-values. Nb: two estimates were obtained in the BARI group and averaged to
26 display only one representative value on the plot. Actual row output was 68% (95%CI [55% ;
27 80%]) (BARI vs TNFi model) or 62% (95% CI [54% ; 70%]) (BARI vs OMA model) for LDA, and
28 23% (95% CI [14% ; 31%]) (BARI vs TNFi model) or 17 % (95% CI [10% ; 24%]) (BARI vs OMA
29 model) for REM.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



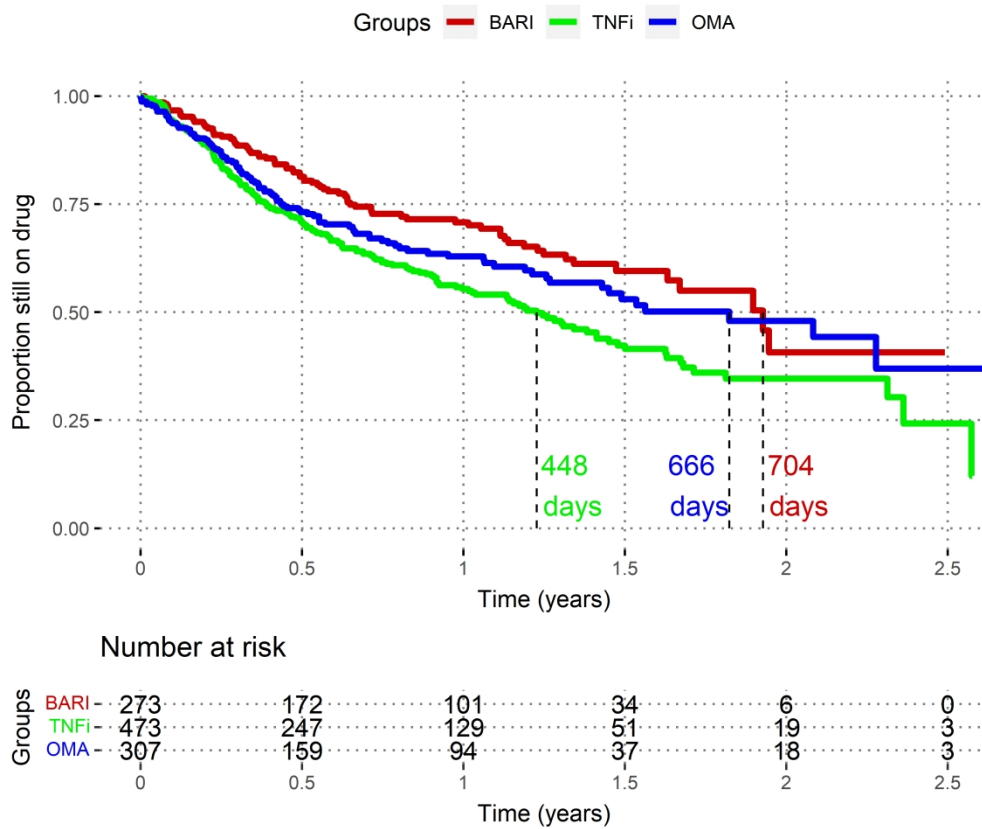


Figure 2 - Non-adjusted time to drug discontinuation analyses (Kaplan-Meier), SCQM, 2017-2020. These "survival curves" represent the drug maintenance after initiation, as the estimated proportion of patients still on therapy, by treatment group. Death and loss to follow-up were censored. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs. Log-Rank BARI vs TNFi : $p < 0.001$. Log-Rank BARI vs OMA: $p = 0.11$.

451x386mm (197 x 197 DPI)

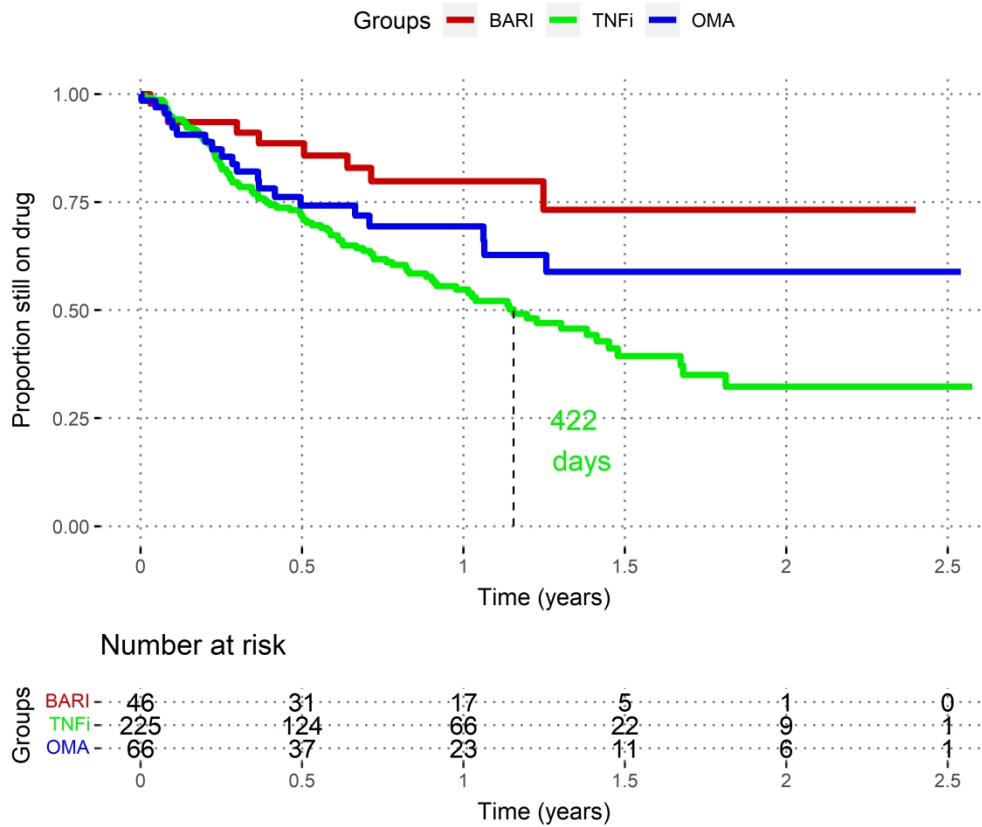


Figure 3 - Unadjusted time to drug discontinuation in b/tsDMARD-naïve patients, SCQM, 2017-2020. These Kaplan-Meier curves represent the crude “survival” of drug prescription, by treatment group. Death and loss to follow-up are censored. BARI = baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs. Log-Rank BARI vs TNFi : $p = 0.003$. Log-Rank BARI vs OMA : $p = 0.15$.

451x386mm (197 x 197 DPI)

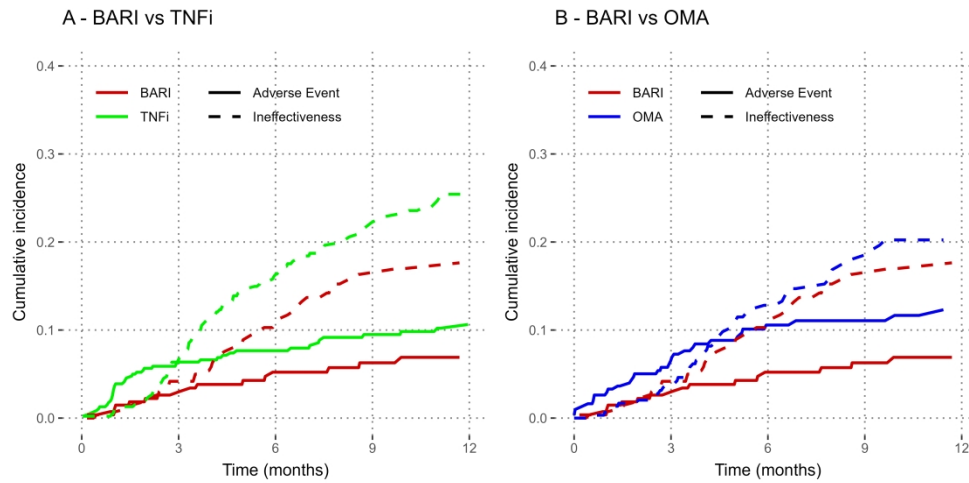


Figure 4 - Cumulative incidence of drug discontinuation by stop reason and by type of treatment, SCQM, 2017-2020.

This figure represents the unadjusted cumulative incidence of drug discontinuation, by group and by reason of discontinuation. BARI = baricitinib. TNFi = Tumor Necrosis Factor. OMA = Other mode of Action bDMARDs.

644x322mm (197 x 197 DPI)

REM and LDA rates
by type of treatment
(CARRAC)

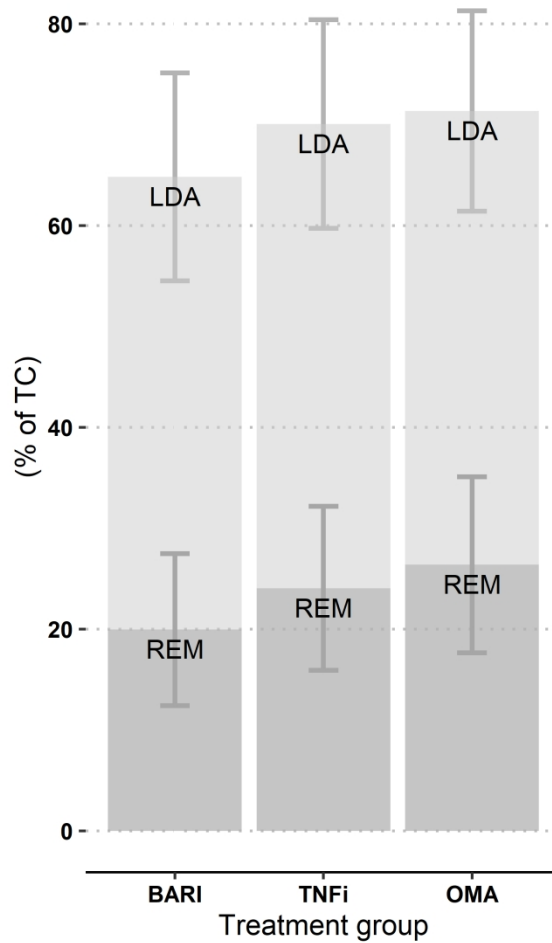


Figure 5 - Estimated response rates at 12-months (CARRAC), SCQM, 2017-2020. BARI = baricitinib. TNFi = Tumor Necrosis Factors Inhibitors. OMA = Other Mode of Action bDMARDs. LDA = Low Disease Activity (i.e. CDAI score <= 10), in light grey. REM = Remission (i.e. CDAI score <= 2.8), in dark grey. 95% CI are represented. This method does not allow computing p-values. Nb: two estimates were obtained in the BARI group and averaged to display only one representative value on the plot. Actual row output was 68% (95%CI [55% ; 80%]) (BARI vs TNFi model) or 62% (95% CI [54% ; 70%]) (BARI vs OMA model) for LDA, and 23% (95% CI [14% ; 31%]) (BARI vs TNFi model) or 17 % (95% CI [10% ; 24%]) (BARI vs OMA model) for REM.

257x451mm (197 x 197 DPI)

SUPPLEMENTARY DATA

Notice on TC duration

Due to frequent changes in medication and short study period, it has to be underlined that the median duration of a TC approximates 200 days. The proportion of TC with follow-up data of at least one year is 37% for BARI, 27% for TNFi and 31% for OMA (Figure S1) - i.e. most TC were started less than 12 months before the date of data extraction.

Notice this % is different from the % of patient still under therapy that we estimate using Kaplan-Meier or Cox model. Indeed, the latter includes a censoring of the lost-to-follow-up patients, hence the denominator is different. As a consequence, this does not contradict the reported “median prescription survival timey”, for instance of 704 days for BARI TCs. The latter is the output estimated by the Kaplan-Meier model, taking censoring into account; it does not imply that actual observations in the dataset have this duration.

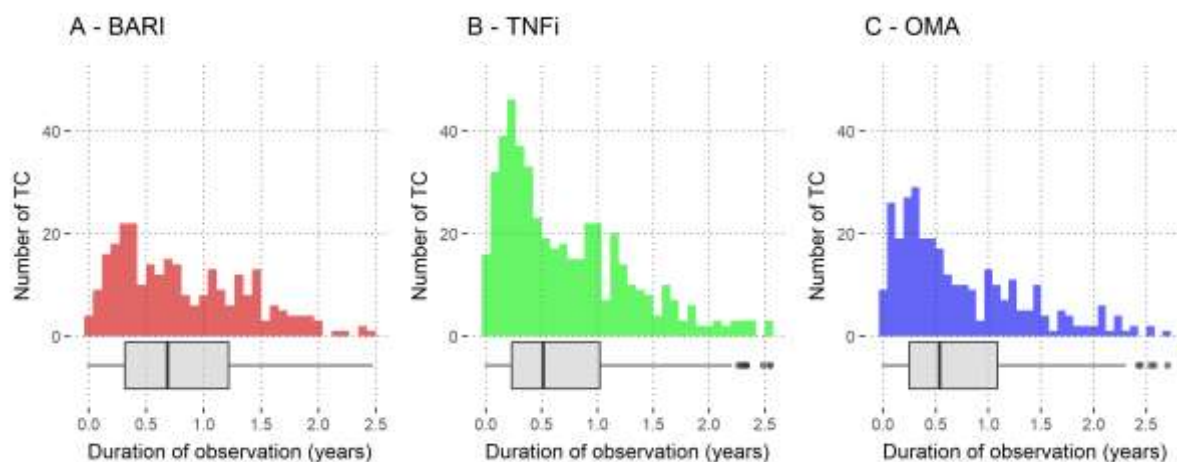


Figure S1: Distribution of the observation time for included TCs, per group, SCQM, 2017-2020.

Most of the treatment courses have an actual duration and/or follow-up period of less than one year. TC = Treatment Course. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs.

Variable definitions

Below we give additional detail about included covariates:

Age: age in years, at TC initiation. Continuous variable.

Gender: male or female. Categorical variable.

BMI: BMI at TC initiation. Continuous variable.

CDAI score: CDAI score at TC initiation. Continuous variable. If missing, imputed according to procedure described in methods section.

Disease duration: time interval between RA diagnosis date and TC initiation date. Continuous variable, expressed in years, but used in decades in models.

Smoking status: smoking status at TC initiation. Categorical variable (current-, former-, never-smoker).

Concomitant csDMARD: yes/no variable. A concomitant csDMARD was defined as csDMARD prescription ongoing for at least 40% of the duration of the TC. Otherwise, the TC was categorized as monotherapy. csDMARDs included: methotrexate, sulfasalazin, leflunomide, azathioprine and hydroxy-chloroquine, alone or in combination.

Concomitant glucocorticoid: Yes/no variable. Concomitant glucocorticoid usage was defined as having at least one active prescription of glucocorticoid, at any dose, at any timepoint of the TC.

Line of therapy: strictly speaking, this categorical variable is displaying: [*number of previous TC ever* + 1]. 4 or more has been grouped in the same category. Hence, it is considering all data of the SCQM registry, i.e. TCs initiated before our study period are also accounted for as previous therapies.

Seropositivity: yes/no variable. Seropositivity is defined as positivity for anti-citrullinated peptide antibodies and/or rheumatoid factor.

Time to all cause discontinuation

Cox model output

Table S1 contains the complete output of the two adjusted Cox models used in the main time-to-drug discontinuation analysis.

Table S1: Hazard ratio of drug discontinuation, Cox models, SCQM-RA registry, 2017-2020.						
	BARI vs TNFi			BARI vs OMA		
	Hazard ratio	95% CI	p	Hazard ratio	95% CI	p
TNFi (vs baricitinib)	1.76	1.32-2.35	<0.001	-	-	-
OMA (vs baricitinib)	-	-	-	1.27	0.93-1.72	0.13
Adjusting variables:						
Age (decades)	1.03	0.92-1.14	0.61	0.98	0.86-1.10	0.69
BMI	1.01	0.98-1.04	0.51	0.98	0.94-1.02	0.31
TC with csDMARD	0.84	0.66-1.09	0.19	1.22	0.90-1.67	0.20
Glucocorticoid usage	1.29	0.93-1.79	0.12	1.86	1.32-2.61	<0.001
CDAI score	1.40	1.26-1.56	<0.001	1.15	1.03-1.28	0.01
Disease duration (decades)	0.95	0.81-1.10	0.46	0.85	0.70-1.03	0.10
Current smoker (vs non-smoker)	1.20	0.86-1.68	0.28	1.09	0.73-1.64	0.66
Ever smoker (vs non-smoker)	1.10	0.79-1.52	0.57	1.38	0.95-2.00	0.09
2nd line therapy (vs 1st)	1.11	0.80-1.53	0.52	1.37	0.81-2.33	0.24
3rd line therapy (vs 1st)	0.98	0.64-1.51	0.93	1.56	0.92-2.64	0.10
4th or later line (vs 1st)	1.06	0.75-1.51	0.73	1.57	0.93-2.63	0.09
Female gender	1.05	0.78-1.42	0.74	1.16	0.81-1.67	0.41
Seropositivity	0.77	0.59-1.01	0.055	0.94	0.67-1.31	0.71
Table S1: BARI = baricitinib. TNFi = TNF inhibitors. OMA = Other Mode of Action drugs. CI = Confidence Interval. BMI = Body Mass Index. CDAI = Clinical Disease Activity Index.						

Figure S2 below gives the exact same information as Table S1:

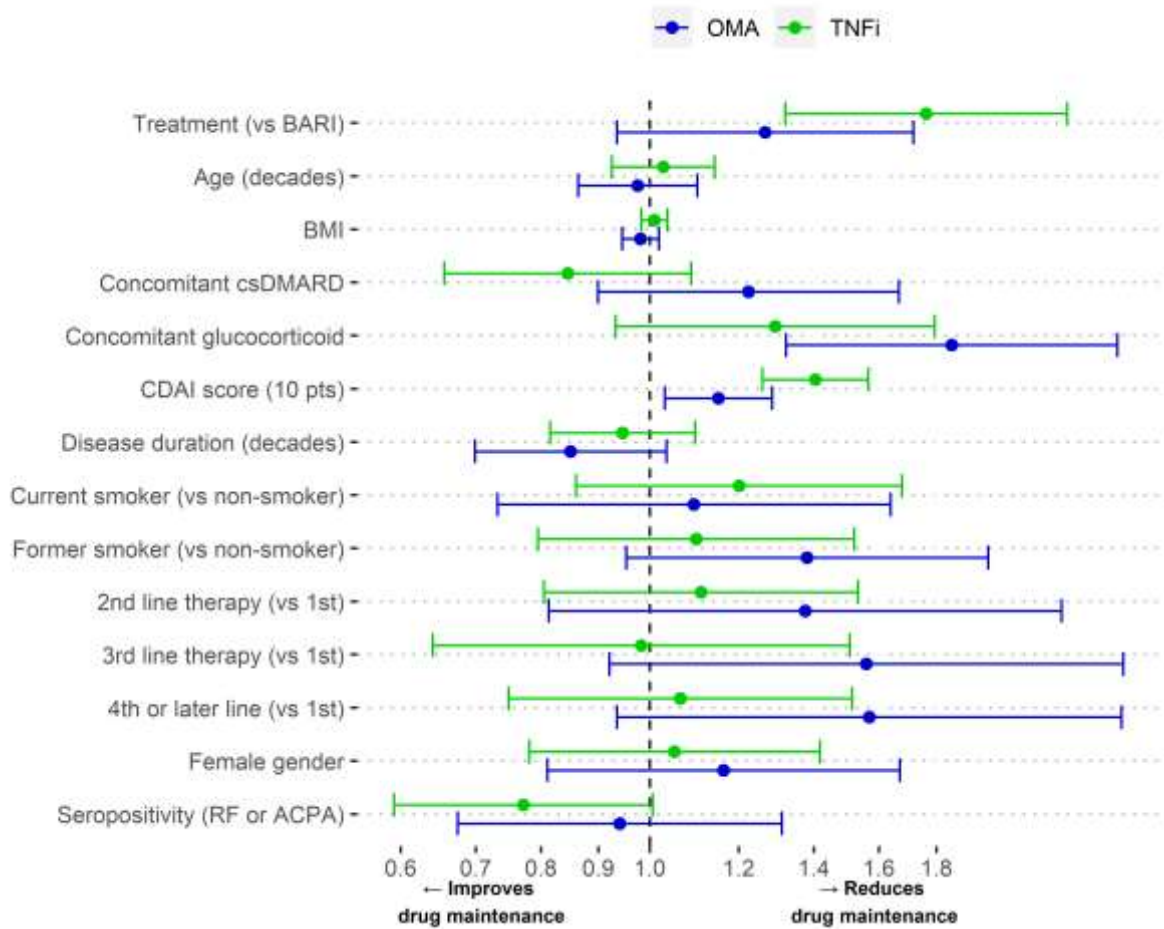


Figure S2: Hazard ratio of drug discontinuation (95% CI).

BARI = Baricitinib. TNFi = Tumor Necrosis Factors Inhibitors. OMA = Other mode of Action bDMARDs. BMI = Body Mass Index. CDAI = Clinical Disease Activity Index. RF = Rheumatoid Factor. ACPA = Anti-citrullinated Peptides Antibodies.

The corresponding cox-adjusted drug-survival curves are provided below:

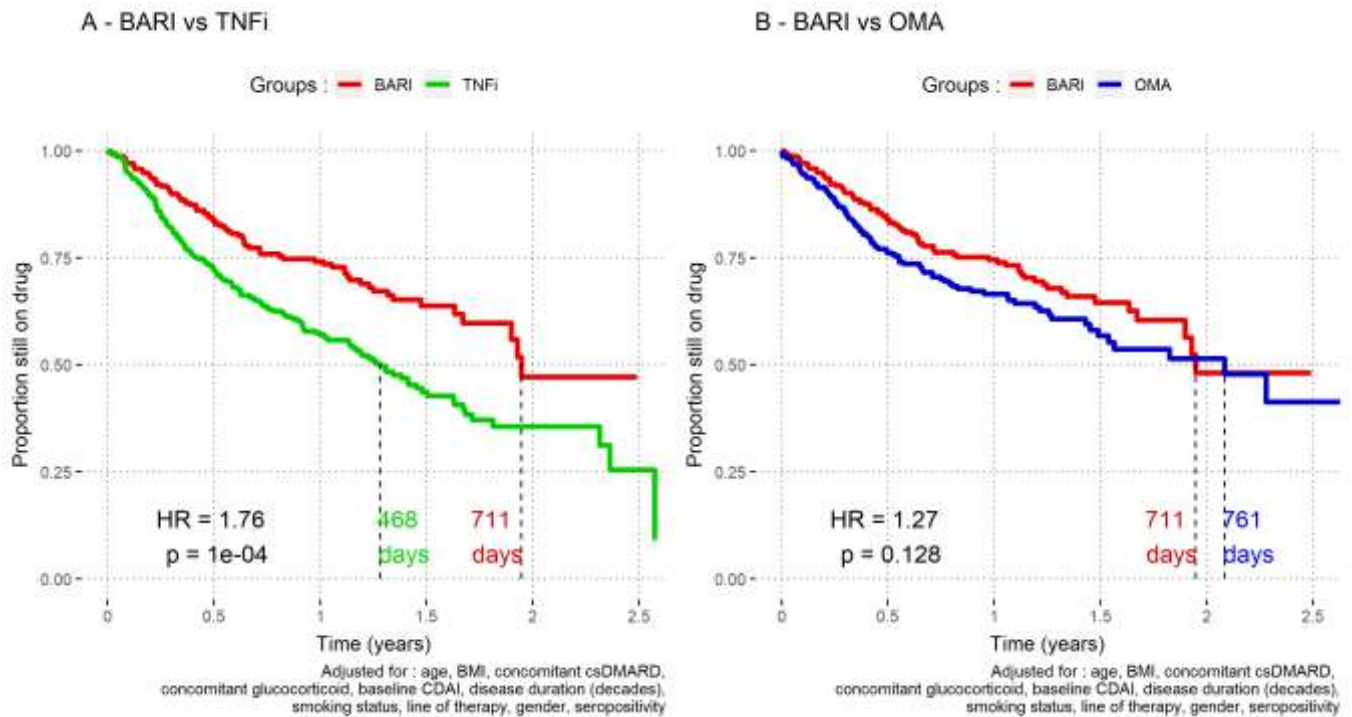


Figure S3: Multivariable Cox model of drug discontinuation by type of treatment, SCQM, 2017-2020.

These curves are merely the visualisation of Cox models presented in Table S1 and Figure S2.

BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action bDMARDs

Models are adjusted for : age, BMI, concomitant csDMARD, concomitant glucocorticoid, baseline CDAI, disease duration, smoking status, line of therapy, gender, seropositivity.

Sensitivity analysis using AIPTW

As a sensitivity analysis, the main time to drug discontinuation was also performed using “augmented inverse probability of treatment weighting” (AIPTW), including the same covariates. In other words, we combined a propensity score using a logistic regression model and an inverse probability weighted Cox regression. We used the *RiskRegression* package in R, to obtain risk ratios.

Figure S4 represents the absolute risk of treatment discontinuation, for all included timepoints. At one year, the adjusted discontinuation risk in BARI was 19 % lower than in TNFi group ($p < 0.001$) (Figure S4 A), with a risk ratio of 1.76 (95% CI [1.19 – 2.34]; $p = 0.009$). Similarly, at one year, the adjusted treatment discontinuation risk in BARI was 8 % lower than in the OMA group ($p = 0.06$) (Figure S4 B), with a risk ratio of 1.28 (95% [0.91 – 1.65]; $p = 0.14$).

Overall, this sensitivity analysis confirms the findings reported in the main body of the article.

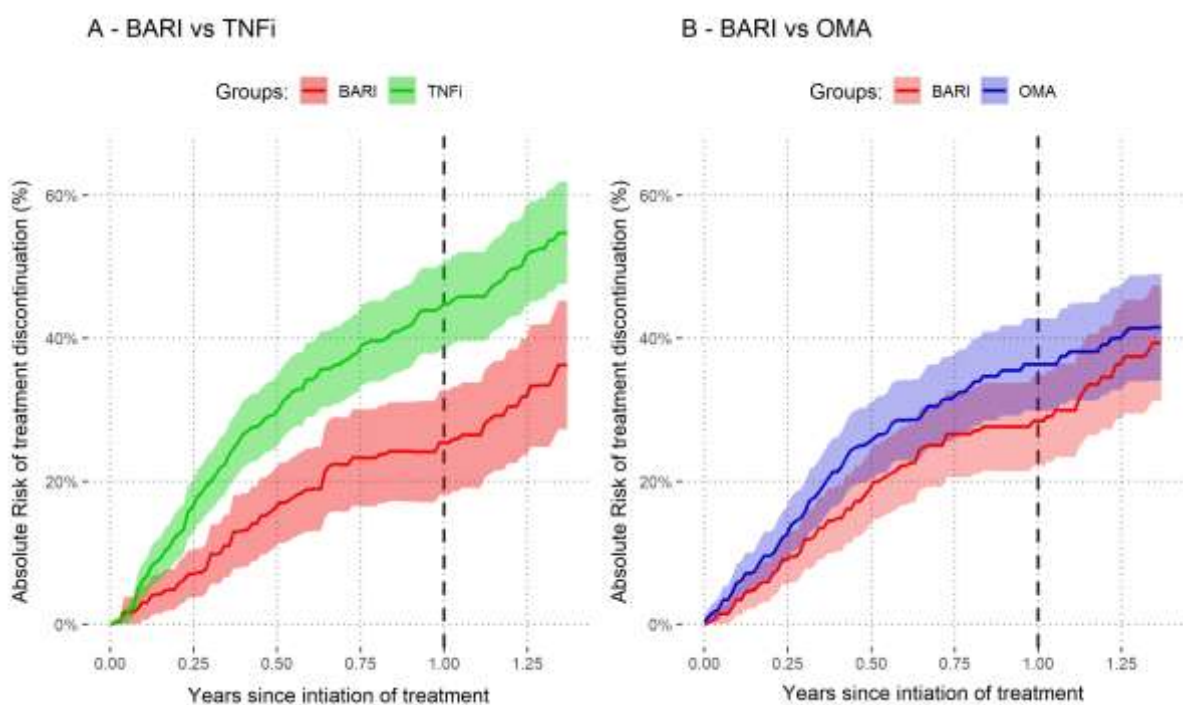


Figure S4: Absolute risk of treatment discontinuation by type of treatment (AIPTW), SCQM, 2017-2020

BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action bDMARDs.

AIPTW = Augmented Inverse Probability of Treatment Weighning. Adjusted for : age, bmi, concomitant csDMARDs, prednisone usage, baseline CDAI, disease duration, smoking status, line of therapy, gender, seropositivity.

Time to all-cause-discontinuation in b/tsDMARD-naïve patients

Table S2: Baseline characteristics of study population, b/tsDMARD-naïve patients, SCQM-RA registry, 2017-2020.

Patient-Variables	BARI (n = 46)		TNFi (n = 225)		OMA (n = 66)		p values
	n %	Miss.	n %	Miss.	n %	Miss.	
Female	70 %	0	71 %	1	73 %	1	0.88
Age (years)	57 (15)	0	52 (14)	1	57 (16)	1	<0.01
Disease duration (years)	6 (6)	1	5 (7)	13	7 (9)	2	0.24
CDAI baseline (raw data)	16 (8)	31	18 (10)	135	18 (14)	42	0.77
CDAI baseline (imputed)	12 (7)	0	14 (9)	0	14 (10)	0	0.61
Obesity (BMI > 30)	11 %	13	13 %	58	5 %	27	0.28
Smoking							
Current	28 %	9	18 %	42	14 %	13	0.18
Former	26 %		24 %		21 %		
Never	26 %		39 %		46 %		
Seropositive (ACPA or RF)	80 %	1	69 %	5	76 %	2	0.20
TC variables							
Dose of BARI (4mg)	83 %	0	-	-	-	-	-
TC duration > 12-months	37 %	0	29 %	0	35 %	0	0.48
Concomitant csDMARD	41 %	0	66 %	0	50 %	0	<0.01
Line of Therapy							
1 st (= bio-naïve)							
2 nd	100 %	0	100 %	0	100 %	0	-
3 rd							
4 th or later							
Previous tsDMARD use (non-BARI)	0 %	0	0 %	0	0 %	0	-
Concomitant glucocorticoid (at any time)	13 %	0	20 %	0	17 %	0	0.50
Mean dose of concomitant glucocorticoid (mg)	1 (4)	0	3 (6)	0	2 (6)	0	0.50

Table S2: BARI = baricitinib. TC = Treatment Courses. CDAI = Clinical Disease Activity Index. TNFi = Tumor Necrosis Factor Inhibitors. OMA = bDMARDs with Other Mode of Action. tsDMARD = targeted synthetic DMARDs. ACPA = Anti Citrullinated Peptide Antibodies. RF = Rheumatoid Factor. Miss. = number of missing values. p-value are computed with either Chi² or ANOVA.

Response rates – raw CDAI data

Figure S6 shows the crude available values for CDAI scores, by type of treatment and time.

Only a minority of CDAI scores were assessed at 0- or 12-month timepoints of TCs (i.e. 680/1053 = 65% were missing for baseline value, and 908/1053 = 86% were missing for exact 12-month value). Future research would certainly benefit having CDAI scores assessed at regular and homogenous time-intervals, based on the initiation date of biological therapies.

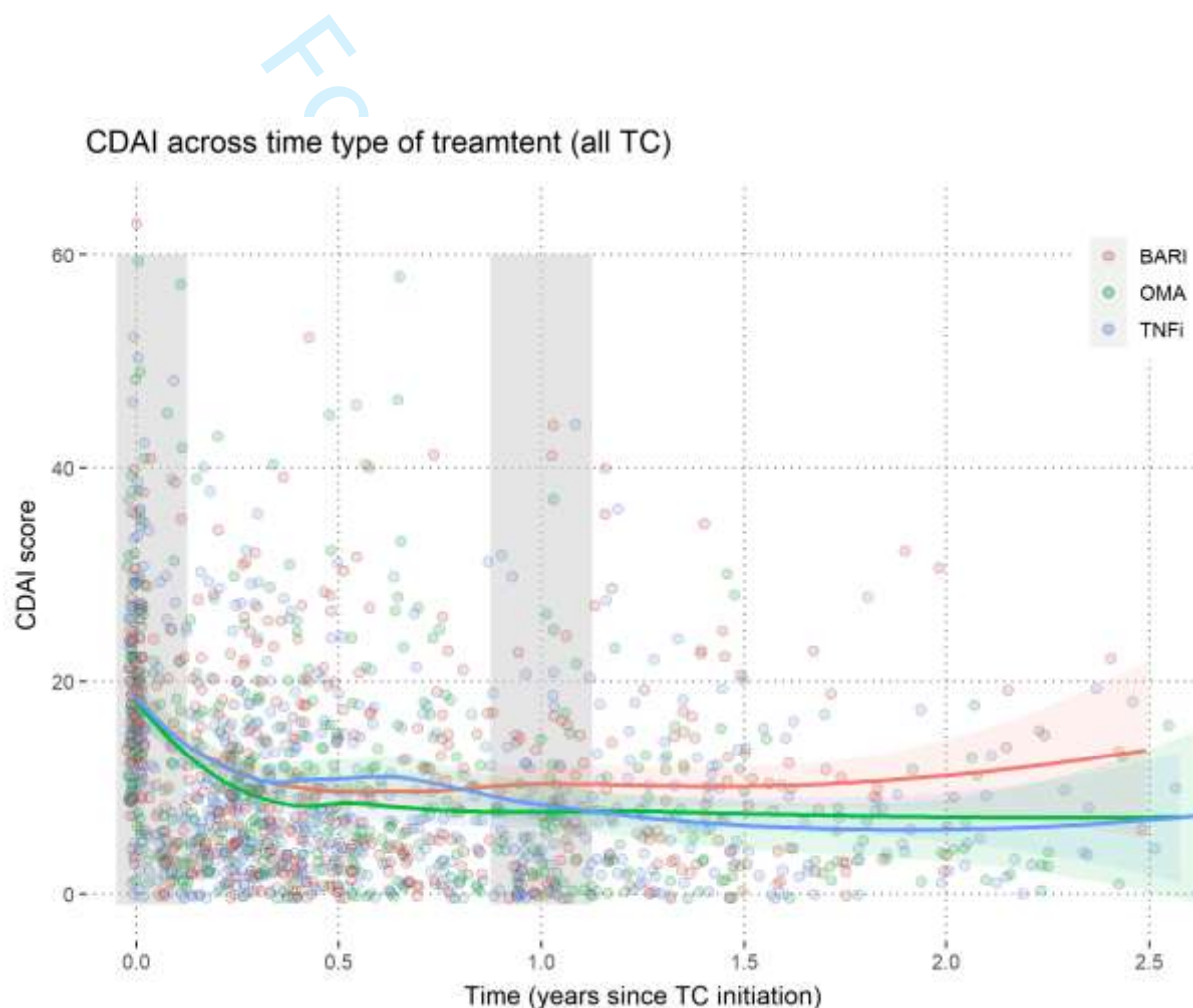


Figure S5: CDAI across time by type of treatment, raw data, SCQM, 2017-2020.

Only a minority of CDAI score were obtained sharp at 0 or 12 months of TCs. CDAI = Clinical Disease Activity Index. TC = Treatment course. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action b/tsDMARDs.

Study size

Based on estimates from similar analyses with tofacitinib (TOFA) performed in this registry, we calculated the number of patients that would be needed to detect a significant decrease in time to all cause-discontinuation of treatment (hazard ratio) between treatment groups using the method described by Schoenfeld and Richter. We assumed a statistical power of 80%, a type I error probability of 0.05, a median BARI retention of 30 months, the inclusion of 3 patients on TNFi for every patient on BARI, an accrual time of 2 years, and additional follow-up of 6 months. We display below the sample size for the BARI group for a range of possible effect sizes (“hazard ratio” between 1.1 and 1.8).

If the true hazard ratio is similar to the one found with TOFA compared to TNFi after a single TNFi failure (HR:1.68) 14, we will need to study 149 patients on BARI and 447 patients on TNFi to be able to reject the null hypothesis that the experimental and control curves are equal with probability (power) of 80%. Pragmatically, we propose to start the analysis of the data only once at least 200 patients on BARI have been included and followed for an average of at least 18 months.

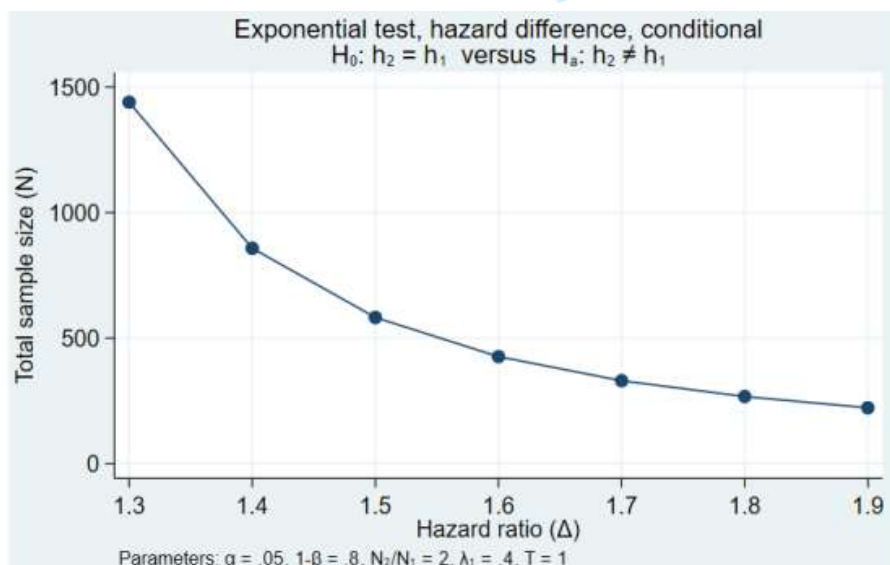


Figure 1: Estimated total sample size for two-sample comparison of survivor functions

1 - SURVIVAL ANALYSIS

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)
library(gridExtra)
library(grid)
library(cmprsk)
library(mstate)
library(cobalt)

rm(list = ls()) # To select all loaded objects and delete them
setwd(dirname(rstudioapi::getActiveDocumentContext())$path) # setting
up working directory in the location of the .Rmd file

load("./1_datamanaged_files/datamanaged.Rdata") # loading data.managed
data

Loading fonctions

# home-made function to force writing with two decimals
formattable = function(nbr){return(formatC(nbr,format = "f",digits =
nombreadpresvirgule))}
nombreadpresvirgule <- 2

# Home-made Fonction to write the p value (by denis)
writepvalue = function(pvalue) {
  if (is.na(pvalue)) {result <- NA} else {
    if(pvalue < 0.001) {
```

```

1
2
3     result <- "<0.001"
4   } else if (pvalue <0.01) {
5     result <- formatC( pvalue ,format = "f",digits = 3)
6   }
7   else {
8     (result <- formatC( pvalue ,format = "f",digits = 2) )
9     i = 1
10    while(result == 0.05) {
11      result <- formatC( pvalue ,format = "f",digits = 2 + i)
12      i = i + 1
13    }
14  }
15  return(result)
16 }
17 }
18 options(scipen = 999)
19
20
21

```

Mini Exploration

```

22 uniqueN(BARI_DATA[]$patient_id) # number of patients (< than number of
23 TC)
24 uniqueN(BARI_DATA[]$TC_id) # unumber of TC
25
26

```

```

27 plot <- qplot(x = BARI_DATA[]$time_on_drug)+
28   geom_vline(xintercept = 365, size = 1.2, alpha = 0.5)+
29   geom_text(aes(x = 365 + 40, label="1 Year", y=20), colour="white",
30   angle=0)+
31
32   geom_vline(xintercept = 2*365, size = 1.2, alpha = 0.5)+
33   geom_text(aes(x = 2*365 + 40, label="2 Year", y=20), colour="black",
34   angle=0)+
35
36   geom_vline(xintercept = mean(BARI_DATA$time_on_drug), color = "red",
37   size = 1.2)+
38   geom_text(aes(x = mean(BARI_DATA$time_on_drug) + 40, label="Mean",
39   y=20), colour="red", angle=0)+
40
41   geom_vline(xintercept = median(BARI_DATA$time_on_drug), color =
42   "green", size = 1.2)+
43   geom_text(aes(x = median(BARI_DATA$time_on_drug) - 40,
44   label="Median", y=20), colour="green", angle=0)+
45
46   labs(x = "Duration of TC (days)", y = "Number of TC", title =
47   "Repartition of the duration of included TC (all groups)")+
48
49   theme_pubclean()
50
51
52
53 plot
54
55 mean(BARI_DATA$time_on_drug)
56
57
58
59
60

```



```

1
2
3 median(BARI_DATA$time_on_drug)
4
5 # Nb : Research protocol said we wanted a follow-up duration of
6 "average of 18 months"
7 mean(BARI_DATA[cohort == "BARI"]$time_on_drug)/30
8 mean(BARI_DATA[cohort == "TNFi"]$time_on_drug)/30
9 mean(BARI_DATA[cohort == "OMA"]$time_on_drug)/30 # looks more like 9
10 months..
11
12
13 # ok, 24-month follow-up will be complicated
14 uniqueN(BARI_DATA[time_on_drug > 2*365, TC_id]) # number if TC with
15 duration > 24 months
16
17
18

```

1. [0] Table 1 BARI vs TNFi et OMA bDMARDs

Common table with all the data

Showing NA to have complete counts and accurate % in each category

```

21
22
23 BARI_DATA[,time_on_drugDiff0 := as.numeric(time_on_drug > 0)] # time
24 in drug < 0
25
26 BARI_DATA[,time_on_drug365 := as.numeric(time_on_drug > 365.25)]
27
28

```

```

29 myVars2 <- c("gender", "age_base", "disease_duration_base_years",
30 "CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
31 "seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
32 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
33 "PREDNISON_STEROID_dose", "dose", "initiation_year",
34 "time_on_drug", "HAQ_score_base")
35

```

```

36 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
37 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
38 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
39 "seropositivity_base")
40

```

```

41 nonnormalVars2 <- c()
42 tab2 <- CreateTableOne(vars = myVars2, data = BARI_DATA, factorVars =
43 catVars2, strata = "cohort", test = F, includeNA = T)
44 tablexp2 <- print(tab2, nonnormal= nonnormalVars2, catDigits = 1,
45 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
46
47

```

saving table 1 NA

```

48
49 write.xlsx(tablexp2, file =
50 "./3_clean_output/BARI_3_groups_table1_NA.xlsx")
51
52

```

Without NA to obtain adequate p values

```

53
54 BARI_DATA[,time_on_drugDiff0 := as.numeric(time_on_drug > 0)] # time
55 in drug < 0
56
57
58
59
60

```

```

1
2
3 BARI_DATA[,time_on_drug365 := as.numeric(time_on_drug > 365.25)]
4
5 myVars2 <- c("gender", "age_base", "disease_duration_base_years",
6 "CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
7 "seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
8 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
9 "PREDNISON_STEROID_dose", "dose", "initiation_year",
10 "time_on_drug", "HAQ_score_base")
11
12
13 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
14 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
15 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
16 "seropositivity_base")
17
18 nonnormalVars2 <- c()
19 tab2 <- CreateTableOne(vars = myVars2, data = BARI_DATA, factorVars =
20 catVars2, strata = "cohort", test = T, includeNA = F)
21 tablexp2 <- print(tab2, nonnormal= nonnormalVars2, catDigits = 1,
22 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
23
24

```

Saving table 1

```

25
26 write.xlsx(tablexp2, file =
27 "./3_clean_output/BARI_3_groups_table1.xlsx")
28

```

But BMJ-Open reviewer 2 aksed for p-values in Table 1 that account for patients providing multiple TCs. Here is how to proceed (it is a bit less conservative):

```

29
30
31
32 library(lme4)
33 library(lmerTest)
34
35
36 # Two glmer() models have to be compared, to assess the impact of
37 grouping, for each baseline variable).
38
39 # gender
40 gender.tab1 <- glmer(gender ~ cohort + (1|patient_id), data =
41 BARI_DATA , family = "binomial")
42 gender.null <- glmer(gender ~ (1|patient_id), data = BARI_DATA ,
43 family = "binomial")
44 anova(gender.tab1, gender.null)
45
46
47 # Age_base
48 age_base.tab1 <- lmer(age_base ~ cohort + (1|patient_id), data =
49 BARI_DATA )
50 age_base.null <- lmer(age_base ~ (1|patient_id), data = BARI_DATA)
51 anova(age_base.tab1, age_base.null)
52
53 # Disease duration
54 disease_duration_base_years.tab1 <- lmer(disease_duration_base_years ~
55 cohort + (1|patient_id), data = BARI_DATA )
56
57
58
59

```

```
1
2
3 disease_duration_base_years.null <- lmer(disease_duration_base_years ~
4 (1|patient_id), data = BARI_DATA)
5 anova(disease_duration_base_years.tab1,
6 disease_duration_base_years.null)
7
8
9
10 # CDAI raw
11 CDAI0_raw.tab1 <- lmer(CDAI0_raw ~ cohort + (1|patient_id), data =
12 BARI_DATA )
13 CDAI0_raw.null <- lmer(CDAI0_raw ~ (1|patient_id), data = BARI_DATA)
14 anova(CDAI0_raw.tab1, CDAI0_raw.null)
15
16
17 # CDAI (imputed)
18 CDAI0.tab1 <- lmer(CDAI0 ~ cohort + (1|patient_id), data = BARI_DATA )
19 CDAI0.null <- lmer(CDAI0 ~ (1|patient_id), data = BARI_DATA)
20 anova(CDAI0.tab1, CDAI0.null)
21
22
23
24 # obesity
25 obese_base.tab1 <- glmer(obese_base ~ cohort + (1|patient_id), data =
26 BARI_DATA , family = "binomial")
27 obese_base.null <- glmer(obese_base ~ (1|patient_id), data = BARI_DATA
28 , family = "binomial")
29 anova(obese_base.tab1, obese_base.null)
30
31
32
33 # smoker base - 1st level vs second level
34 smoker_base.tab1 <- glmer(smoker_base ~ cohort + (1|patient_id), data
35 = BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "FORMER_SMOKER")] ,
36 family = "binomial")
37 smoker_base.null <- glmer(smoker_base ~ (1|patient_id), data =
38 BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "FORMER_SMOKER")] ,
39 family = "binomial")
40 anova(smoker_base.tab1, smoker_base.null)
41
42
43
44 # smoker base - 2nd level versus third
45 smoker_base.tab1 <- glmer(smoker_base ~ cohort + (1|patient_id), data
46 = BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "NEVER_SMOKER")] ,
47 family = "binomial")
48 smoker_base.null <- glmer(smoker_base ~ (1|patient_id), data =
49 BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "NEVER_SMOKER")] ,
50 family = "binomial")
51 anova(smoker_base.tab1, smoker_base.null)
52
53
54
55 # seropositivity
56 seropositivity_base.tab1 <- glmer(seropositivity_base ~ cohort + (1|
```

```
1
2
3 patient_id), data = BARI_DATA , family = "binomial")
4 seropositivity_base.null <- glmer(seropositivity_base ~ (1|
5 patient_id), data = BARI_DATA , family = "binomial")
6 anova(seropositivity_base.tab1, seropositivity_base.null)
7
8
9
10 # Concomittant csDMARD
11 TC_with_csDMARD.tab1 <- glmer(TC_with_csDMARD ~ cohort + (1|
12 patient_id), data = BARI_DATA , family = "binomial")
13 TC_with_csDMARD.null <- glmer(TC_with_csDMARD ~ (1|patient_id), data =
14 BARI_DATA , family = "binomial")
15 anova(TC_with_csDMARD.tab1, TC_with_csDMARD.null)
16
17
18
19 # Line of therapy
20 line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
21 (1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "2nd")]
22 , family = "binomial")
23 line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
24 patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "2nd")] ,
25 family = "binomial")
26 anova(line_of_therapy.tab1, line_of_therapy.null)
27
28
29
30 # Line of therapy
31 line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
32 (1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "3rd")]
33 , family = "binomial")
34 line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
35 patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "3rd")] ,
36 family = "binomial")
37 anova(line_of_therapy.tab1, line_of_therapy.null)
38
39
40
41 # Line of therapy
42 line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
43 (1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st",
44 "4th_or_later")] , family = "binomial")
45 line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
46 patient_id), data = BARI_DATA[line_of_therapy %in% c("1st",
47 "4th_or_later")] , family = "binomial")
48 anova(line_of_therapy.tab1, line_of_therapy.null)
49
50
51
52 # N_prev_tsDMARD
53 N_prev_tsDMARD.tab1 <- glmer(as.factor(N_prev_tsDMARD) ~ cohort + (1|
54 patient_id), data = BARI_DATA , family = "binomial")
55 N_prev_tsDMARD.null <- glmer(as.factor(N_prev_tsDMARD) ~ (1|
56 patient_id), data = BARI_DATA , family = "binomial")
57
58
59
```

```

1
2
3 anova(N_prev_tsDMARD.tab1, N_prev_tsDMARD.null)
4
5
6
7 # Concomittant prednisone
8 PREDNISON_STEROID.tab1 <- glmer(PREDNISON_STEROID ~ cohort + (1|
9 patient_id), data = BARI_DATA , family = "binomial")
10 PREDNISON_STEROID.null <- glmer(PREDNISON_STEROID ~ (1|patient_id),
11 data = BARI_DATA , family = "binomial")
12 anova(PREDNISON_STEROID.tab1, PREDNISON_STEROID.null)
13
14
15 # Dose of PREDNISON
16 PREDNISON_STEROID_dose.tab1 <- lmer(PREDNISON_STEROID_dose ~ cohort +
17 (1|patient_id), data = BARI_DATA )
18 PREDNISON_STEROID_dose.null <- lmer(PREDNISON_STEROID_dose ~ (1|
19 patient_id), data = BARI_DATA)
20 anova(PREDNISON_STEROID_dose.tab1, PREDNISON_STEROID_dose.null)
21
22
23 Other various computations for Table 1 --
24 uniqueN(BARI_DATA$patient_id)
25
26 mean(BARI_DATA[, time_on_drug])
27
28 median(BARI_DATA[cohort == "Bari" , time_on_drug])
29 median(BARI_DATA[cohort == "OMA" , time_on_drug])
30 median(BARI_DATA[cohort == "TNFi" , time_on_drug])
31
32 median(BARI_DATA[cohort == "OMA"]$time_on_drug)
33
34
35 mean(BARI_DATA[,disease_duration_base_years], na.rm = T)
36
37 table(is.na(BARI_DATA$CDAI0_raw), BARI_DATA$cohort) # number of
38 missing CDAI0_raw...
39 table(is.na(BARI_DATA$CDAI0), BARI_DATA$cohort) # number of missing
40 CDAI0... (after imputation)
41
42
43 table(is.na(BARI_DATA$CDAI12_raw), BARI_DATA$cohort) # number of
44 missing CDAI12_raw...
45 table(is.na(BARI_DATA$CDAI12), BARI_DATA$cohort) # number of missing
46 CDAI12 after imputation
47
48 hist(BARI_DATA$CDAI0_raw)
49 hist(BARI_DATA$CDAI0)
50
51 summary(BARI_DATA[cohort=="BARI", c("gender",
52 "age_base","disease_duration_base_years", "CDAI0_raw", "CDAI0",
53 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
54 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
55 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
56
57
58
59
60

```

```

1
2
3 "initiation_year", "time_on_drug", "HAQ_score_base")) # to see NA
4 values for all variables
5

```

```

6 summary(BARI_DATA[cohort=="TNFi", c("gender",
7 "age_base", "disease_duration_base_years", "CDAI0_raw", "CDAI0",
8 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
9 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
10 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
11 "initiation_year", "time_on_drug", "HAQ_score_base")) # to see NA
12 values for all variables
13

```

```

14
15 summary(BARI_DATA[cohort=="OMA", c("gender",
16 "age_base", "disease_duration_base_years", "CDAI0_raw", "CDAI0",
17 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
18 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
19 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
20 "initiation_year", "time_on_drug", "HAQ_score_base")) # to see NA
21 values for all variables
22

```

```

23
24 table(is.na(BARI_DATA$disease_duration_base_years), BARI_DATA$cohort)
25 # number of missing disease duration...
26

```

```

27 table(is.na(BARI_DATA$age_base), BARI_DATA$cohort) # number of steroid
28 doses missing
29

```

```

30 table(is.na(BARI_DATA$PREDNISON_STEROID_dose), BARI_DATA$cohort) #
31 number of missing baseline steroids
32

```

Imputation using MICE -- BARI vs TNFi et OMA bDMARDs --

Common imputation step with all data

```

33
34
35
36
37 BARI <- BARI_DATA[,c("TC_id", "patient_id", "stop_DMARD",
38 "stop_reasons", "age_base", "concomitant_csDMARD",
39 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
40 "CDAI0", "disease_duration_base_years", "time_on_drug", "bmi_base",
41 "smoker_base", "line_of_therapy", "obese_base", "gender", "cohort",
42 "adverse_event_reported", "seropositivity_base")] # choose variables
43 of interest
44

```

```

45
46 BARI$smoker_base <- as.factor(BARI$smoker_base) # put labels as factor
47 BARI$line_of_therapy <- as.factor(BARI$line_of_therapy)
48 BARI$gender <- as.factor(BARI$gender)
49 BARI$concomitant_csDMARD <- as.factor(BARI$concomitant_csDMARD)
50 BARI$PREDNISON_STEROID <- as.factor(BARI$PREDNISON_STEROID)
51 BARI$cohort <- as.factor(BARI$cohort)
52

```

```

53
54 # Imputation
55

```

```

56 if(!file.exists("./2_cached_files/imputed_data")){ # to avoid re-
57
58
59

```

```

1
2
3 computing if already done
4   imputed_data <- mice(BARI, m=50, method="pmm", maxit=25, seed=500)
5   save(imputed_data, file = "./2_cached_files/imputed_data")
6 } else {
7   load("./2_cached_files/imputed_data")
8 }
9
10 # Subsettings
11
12 BARI1 <- BARI[cohort %in% c("BARI", "TNFi")] # creating subset for
13 BARI vs TNFi comparasion
14 BARI1[,cohort := as.factor(as.character(cohort))]
15
16
17 imputed_data1 <- complete(imputed_data,"long", include=T) # to put in
18 long format and categorize variables
19 imputed_data1 <- imputed_data1[imputed_data1$cohort %in% c("BARI",
20 "TNFi"),] # to keep only BARI and TNFi rows
21 imputed_data1$cohort <- as.factor(as.character(imputed_data1$cohort))
22 imputed_data1 <- as.mids(imputed_data1) # re concatenating in previous
23 format, to use fit.mult.impute
24
25
26
27 BARI2 <- BARI[cohort %in% c("BARI", "OMA")] # creating subset for BARI
28 vs OMA comparasion
29 BARI2[,cohort := as.factor(as.character(cohort))]
30
31
32 imputed_data2 <- complete(imputed_data,"long", include=T) # to put in
33 long format and categorize variables
34 imputed_data2 <- imputed_data2[imputed_data2$cohort %in% c("BARI",
35 "OMA"),] # to keep only BARI and OMA rows
36 imputed_data2$cohort <- as.factor(as.character(imputed_data2$cohort))
37 imputed_data2 <- as.mids(imputed_data2) # re concatenating in previous
38 format, to use fit.mult.impute
39
40

```

1. [1] SURVIVAL ANALYSIS (drug discontinuation)

Exploration

```

44 table(BARI_DATA$cohort, BARI_DATA$stop_DMARD)
45 table(BARI_DATA$cohort, BARI_DATA$stop_reasons)
46
47

```

Checking adequacy of COX models --

For BARI vs TNFi

```

51 # categorization for linearity checking
52 test1 <- complete(imputed_data1,"long", include=T) # to put in long
53 format and categorize variables
54
55
56
57
58
59

```

```

1
2
3 test1$agecat <- cut(test1$age_base, 4)
4 test1$bmicat <- cut(test1$bmi_base, 4)
5 test1$cdaicat <- cut(test1$CDAI0, 4)
6 test1$duracat <- cut(test1$disease_duration_base_years, 4)
7
8
9 test1 <- as.mids(test1, .imp=1, .id=2) # re concatenating in previous
10 format, to use fit.mult.impute
11
12 # linearity checking
13
14 BARI1.adj.mi.test <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
15 as.factor(cohort)+
16                                     I(agecat)+
17                                     I(bmicat)+
18                                     TC_with_csDMARD+
19                                     PREDNISON_STEROID+
20                                     I(cdaicat)+
21                                     I(duracat)+
22                                     C(smoker_base, base=3)+
23                                     line_of_therapy+
24                                     gender+
25                                     seropositivity_base+
26                                     cluster(patient_id),
27                                     fitter = coxph, xtrans = test1,
28
29 data = BARI1)
30
31 summary(BARI1.adj.mi.test)
32 rm(BARI1.adj.mi.test)
33
34 # Log-linearity of coefficients ?
35
36 # Coefs age are between 0.15 and 0.25, let's assume it's ok
37 # Hum bmi coefs are not so log-linear, rather close to 0
38 # For CDAI also
39 # Looks ok for disease_duration_base_years.
40
41 # --> Let's keep all variable in the continuous format
42
43 # Proportionality test of hazards on raw data
44
45
46 test1ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
47 as.factor(cohort)+
48                                     cluster(patient_id),
49                                     data= BARI1)
50
51 cox.zph(test1ph) # it's ok
52
53 # Hazard proportionality test on imputed data sets
54
55 test1 <- complete(test1,"long",include=T) # To reset the charges to
56
57
58
59
60

```



```
1
2
3 long format
4 test1 <- test1[test1$.imp==1 | test1$.imp==2 | test1$.imp==3 |
5 test1$.imp==4 | test1$.imp==5 ,] # To select only 5 data sets
6
7 test1ph.adj.mi <- coxph(Surv(time = time_on_drug, event = stop_DMARD)
8 ~ as.factor(cohort)+
9
10         I(age_base/10)+
11         bmi_base+
12         TC_with_csDMARD+
13         PREDNISON_STEROID+
14         CDAI0+
15         I(disease_duration_base_years/10)+
16         C(smoker_base, base=3)+
17         line_of_therapy+
18         gender+
19         seropositivity_base+
20         cluster(patient_id),
21         data = test1)
22
23
24 cox.zph(test1ph.adj.mi)
25
26 schonfeldall <- cox.zph(test1ph.adj.mi) # Test cox.zph may not be ok,
27 but it's because of the multiple imputation (often the case with a lot
28 of data)
29 for (i in 1:11){
30     plot(schonfeldall[i]) # so we should go to a visual testing --> ok
31 }
32
33 rm(schonfeldall, test1ph.adj.mi, test1)
34
35 For BARI vs OMA
36
37 # categorization for linearity checking
38 test2 <- complete(imputed_data2,"long", include=T) # to put in long
39 format and categorize variables
40
41 test2$agecat <- cut(test2$age_base, 4)
42 test2$bmicat <- cut(test2$bmi_base, 4)
43 test2$cdai0cat <- cut(test2$CDAI0, 4)
44 test2$duracat <- cut(test2$disease_duration_base_years, 4)
45
46
47 test2 <- as.mids(test2, .imp=1, .id=2) # re concatenating in previous
48 format, to use fit.mult.impute
49
50 # linearity checking
51
52 BARI2.adj.mi.test <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
53 as.factor(cohort)+
54
55         I(agecat)+
56         I(bmicat)+
57
58
59
60
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

TC_with_csDMARD+
PRÉDNISÓN_STEROID+
I(cdaicat)+
I(duracat)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base+
cluster(patient_id),
fitter = coxph, xtrans = test2,
data = BARI2)

summary(BARI2.adj.mi.test)
rm(BARI2.adj.mi.test)

# Log-linearity of coefficients ?

# Coefs age are around -0.4, let's assume it's ok
# Hum bmi coefs are discussable
# For CDAI it's ok
# More or less ok for disease_duration_base_years.

# --> Let's keep all variable in the continuous format

# Proportionality test of hazards on raw data

test2ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
as.factor(cohort)+
cluster(patient_id),
data= BARI2)
cox.zph(test2ph) # it's ok

# Hazard proportionality test on imputed data sets

test2 <- complete(test2,"long",include=T) # To put imputed data in
ling format
test2 <- test2[test2$.imp==1 | test2$.imp==2 | test2$.imp==3 |
test2$.imp==4 | test2$.imp==5 ,] # To select only 5 datasets

test2ph.adj.mi <- coxph(Surv(time = time_on_drug, event = stop_DMARD)
~ as.factor(cohort)+
I(age_base/10)+
bmi_base+
TC_with_csDMARD+
PRÉDNISÓN_STEROID+
CDAI0+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+

```

```

1
2
3         line_of_therapy+
4         gender+
5         seropositivity_base+
6         cluster(patient_id),
7         data=test2)
8
9
10    cox.zph(test2ph.adj.mi)
11
12    schonfeldall <- cox.zph(test2ph.adj.mi) # Test cox.zph may not be ok,
13    but it's because of the multiple imputation (often the case with a lot
14    of data)
15    for (i in 1:11){
16        plot(schonfeldall[i]) # so we should go to a visual testing --> ok
17    }
18
19    rm(schonfeldall, test2ph.adj.mi, test2)

```

BARI vs TNFi --

COX model

Final Cox Model

```

27    BARI1.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
28    cohort +
29
30        I(age_base/10)+
31        bmi_base+
32        TC_with_csDMARD+
33        PREDNISON_STEROID+
34        I(CDAI0/10)+
35        I(disease_duration_base_years/10)+
36        C(smoker_base, base=3)+
37        line_of_therapy+
38        gender+
39        seropositivity_base+
40        cluster(patient_id),
41        fitter = coxph, xtrans = imputed_data1,
42    data = BARI1)
43
44    summary(BARI1.adj.mi)

```

Creation of HR table and p-values

```

48    ploufrows <- names(BARI1.adj.mi$coefficients)
49    ploufcols <- c("HR", "95%CI", "p")
50    coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
51    length(ploufcols))
52    rownames(coxtable) <- ploufrows
53    colnames(coxtable) <- ploufcols
54    plouf <- summary(BARI1.adj.mi)

```

```

1
2
3   for(row in ploufrows)
4   {
5     coxtable[row,"HR"] <-
6     formattable(plouf$coefficients[row,"exp(coef)"])
7     coxtable[row,"95%CI"] <-
8     paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",
9     formattable(plouf$conf.int[row,"upper .95"]))
10    coxtable[row,"p"] <- writepvalue(plouf$coefficients[row,"Pr(>|
11    z|)"])
12  }
13
14
15  write.xlsx(coxtable, file="./3_clean_output/BARI vs TNFi HR.xlsx") #
16  saving excel file
17
18  Forest plot
19  meanall <- summary(BARI1.adj.mi)$coefficients[1:14,"exp(coef)"]
20  lowerall <- summary(BARI1.adj.mi)$conf.int[1:14,"lower .95"]
21  upperall <- summary(BARI1.adj.mi)$conf.int[1:14,"upper .95"]
22  textall <- c("TNFi (vs BARI)", "Age (decades)", "BMI", "Concomitant
23  csDMARD", "Concomitant glucocorticoid", "CDAI score (10 pts)",
24  "Disease duration (decades)", "Current smoker (vs non-smoker)",
25  "Former smoker (vs non-smoker)", "2nd line therapy (vs 1st)", "3rd
26  line therapy (vs 1st)", "4th or later line (vs 1st)", "Female gender",
27  "Seropositivity (RF or ACPA)")
28
29
30  dfall <- data.frame(textall, meanall, lowerall, upperall)
31  dfall$textall <- factor(dfall$textall,
32    levels = textall)
33
34  HR_plot_1 <- ggplot(data=dfall, aes(x=textall, y= meanall, ymin =
35  lowerall, ymax = upperall))+
36
37    geom_pointrange(size=0.5)+
38    geom_errorbar(aes(ymin=lowerall, ymax=upperall),width=0.5)+
39    geom_hline(yintercept =1, linetype=2)+
40
41    xlab('')+ ylab(" ")
42    ggtitle("BARI vs TNFi")+
43
44    scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
45  +
46  facet_wrap(~textall,nrow=16, strip.position= "right", scales =
47  "free_y") +
48
49
50  theme_pubclean()+
51  theme(strip.text.y = element_blank(),
52    strip.background = element_blank(),
53    axis.line.x = element_line(size = 0.5),
54    axis.text = element_text(face = "bold", colour = "black"),
55    legend.position="bottom", plot.margin =
56
57
58
59

```

```
1
2
3 unit(c(1,3,2,1),"lines"))+
4
5     coord_flip()
6
7 HR_plot_1
8
9
10 # adding some manual annotation
11 grid.text("Improves drug maintenance", x = unit(0.3, "npc"), y =
12 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
13 grid.text("Reduces drug maintenance", x = unit(0.87, "npc"), y =
14 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
15
```

Non-adjusted Kaplan-Meier curves

based on mini-tutorial found on datacamp.com/community/tutorials/survival-analysis-R

BARI vs TNFi

```
21 surv_object1 <- Surv(time = BARI1$time_on_drug, event =
22 BARI1$stop_DMARD) # indicate time on drug and stop variable
23 summary(coxph(surv_object1 ~ cohort, data=BARI1))
24 fit1 <- survfit(surv_object1 ~ cohort, data = BARI1) # this function
25 creates the data for Kaplan Meyer
26 fit1
27
28 survplot_1 <- ggsurvplot(fit1, data = BARI1, # plot
29 pval = T,
30 pval.method = TRUE,
31 legend.title = "Groups :",
32 legend.labs = c("Baricitinib", "TNFi"),
33 xlab = "Time (days)",
34 xlim = c(0, 700),
35 censor = FALSE,
36 title = "Non-adjusted model of drug discontinuation by type
37 of treatment",
38 surv.median.line = "v",
39 linetype = 1,
40 size = 1.5,
41 ggtheme = theme_minimal(),
42 #palette = c("grey78", "grey10"),
43 palette = c("red2", "green3"), # specify colors
44 risk.table = T)
45
46 survplot_1
47 summary(fit1, times = 365)
48 summary(fit1, times = 730)
49
```

Saving the plot curv object for Lilly

```
50
51
52 plot_BARI_vs_TNFi_data <- survplot_1$data.survplot
53 write.xlsx(plot_BARI_vs_TNFi_data, file =
54 "./3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_non_adjust
55 ed.xlsx", row.names = F)
56
57
58
59
```

Home-made attempt to obtain adjusted curves based on imputed data

```
1
2
3
4 dummy_cox_impute1 <- mice::complete(imputed_data1, "long", include =
5 T)
6 dummy_cox_impute1 <- dummy_cox_impute1[dummy_cox_impute1$.imp != 0,]
7
8
9 BARI_fit1 <- survfit(coxph(Surv(time = time_on_drug, event =
10 stop_DMARD) ~ cohort+
11 I(age_base/10)+
12 bmi_base+
13 TC_with_csDMARD+
14 PREDNISON_STEROID+
15 CDAI0+
16 I(disease_duration_base_years/10)+
17 C(smoker_base, base=3)+
18 line_of_therapy+
19 gender+
20 seropositivity_base+
21 cluster(patient_id)+
22 strata(cohort),dummy_cox_impute1), data =
23 dummy_cox_impute1)
24
25
26
27 survplot_1_adj <- ggsurvplot(BARI_fit1, data = dummy_cox_impute1,
28 variable = "cohort",
29 xlab = "Time (days)",
30 title = "Multivariable Cox model of drug discontinuation by
31 type of treatment - BARI vs TNFi",
32 legend.title = "Groups :",
33 legend.labs = c("Baricitinib", "TNFi"),
34 censor = FALSE,
35 xlim = c(0, 700),
36 surv.median.line = "v",
37 linetype = 1,
38 size = 1.5,
39 ggtheme = theme_minimal(),
40 # palette = c("grey78", "grey10")
41 palette = c("red2", "green3") # to change colors
42 )
43
44
45 # adding some legends
46 survplot_1_adj <- survplot_1_adj +
47 labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
48 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
49 smoking status, line of therapy, gender, seropositivity")
50
51
52 survplot_1_adj
53 # summary(BARI_fit1) # to see detailed surv probabilities at given
54 timepoints
55
56
57
58
59
60
```

```
1
2
3 summary(BARI_fit1, times = 365)
4 summary(BARI_fit1, times = 730)
5
6 Saving the plot curv object for Lilly
7
8 plot_BARI_vs_TNFi_data_adj <- survplot_1_adj$data.survplot
9 write.xlsx(plot_BARI_vs_TNFi_data_adj, file =
10 ".\3_clean_output\Lilly_curves_excel\plot_BARI_vs_TNFi_data_adj.xlsx",
11 row.names = F)
12
```

Sensitivity analysis with package RiskRegression (AIPTW)

```
13
14 # Rappel: imputed_data1 = BARI vs TNFi
15 #           imputed_data2 = BARI vs OMA
16 library(riskRegression)
17
18 # I select only one imputed dataset. Would be even better to find a
19 way to pool/average the results from the 50 imputed datasets, but it
20 does not seem doable by default
21 test.data <- complete(imputed_data1, 1)
22
23
24 # First, we specify the treatment model (propensity score model)
25 # Logistic regression where the treatment group is the dependent
26 variable.
27
28 m.treatment <- glm(cohort~I(age_base/10)+
29                   bmi_base+
30                   TC_with_csDMARD+
31                   PREDNISON_STEROID+
32                   I(CDAI0/10)+
33                   I(disease_duration_base_years/10)+
34                   C(smoker_base, base=3)+
35                   line_of_therapy+
36                   gender+
37                   seropositivity_base,
38                   data = test.data, family =
39 "binomial" )
40
41
42 # Then we specify both the "event model" and the "censoring model".
43 Both are cox model
44
45 m.event <- coxph(Surv(time_on_drug, stop_DMARD) ~ cohort+
46                 I(age_base/10)+
47                 bmi_base+
48                 TC_with_csDMARD+
49                 PREDNISON_STEROID+
50                 I(CDAI0/10)+
51                 I(disease_duration_base_years/10)+
52                 C(smoker_base, base=3)+
53                 line_of_therapy+
54                 gender+
55
56
57
58
59
60
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

    seropositivity_base,
    data = test.data, x = TRUE, y =
TRUE)

m.censor <- coxph(Surv(time_on_drug, stop_DMARD==0) ~ cohort +
    I(age_base/10)+
    bmi_base+
    TC_with_csDMARD+
    PREDNISON_STEROID+
    I(CDAI0/10)+
    I(disease_duration_base_years/10)+
    C(smoker_base, base=3)+
    line_of_therapy+
    gender+
    seropositivity_base
    , x =TRUE, y = TRUE,
    data = test.data)

# And we measure the average treatment effect using function "ate",
specifying the time at which we want to compute the ATE

out <- ate(event = m.event ,
    treatment = m.treatment,
    censor = m.censor,
    data = test.data,
    cause = 1,
    estimator = "AIPTW",
    times = seq(from = 0, to = 500, by = 5))

dt.out <- as.data.table(out)

Diagnostics asked by Lily statistician

library(cobalt)

# First, the distribution of propensity scores
test.data$pscores <- m.treatment$fitted.values
test.data %>% setDT()

pscore_plot <- ggplot(test.data, aes(x = pscores, color = cohort, fill
= cohort)) +
  geom_density(alpha = .47) +
  xlab("Estimated Probability of being assigned BARI") +
  ylab("Density") +
  theme_minimal()+
  theme(axis.ticks.y = element_blank(),
    panel.grid.minor = element_blank(),
    legend.title = element_blank(),
    text = element_text(size = 16),

```



```
1
2
3       axis.title.x = element_text(hjust = 0.2, size = 16))
4 pscore_plot # overlap
5
6
7
8 ## Computing the weights
9 test.data$weights <- ifelse(test.data$cohort == "TNFi",
10 1/test.data$pscores, 1/(1-test.data$pscores))
11
12 # Selecting only our covariates of interest (the ones in the ps model)
13 COVS <- subset(test.data, select = c(cohort,age_base,
14                                     bmi_base,
15                                     TC_with_csDMARD,
16                                     PREDNISOLONE_STEROID,
17                                     CDAI0,
18                                     disease_duration_base_years,
19                                     smoker_base,
20                                     line_of_therapy,
21                                     gender,
22                                     seropositivity_base))
23
24
25 # To get the SMD & variance ratios before/after weighting
26 # bal.tab(COVS, treat = test.data$cohort, thresholds = 0.1)
27 # bal.tab(COVS, treat = test.data$cohort, weights = test.data$weights,
28 thresholds = 0.1)
29 # bal.tab(COVS, treat = test.data$cohort, v.threshold = 2)
30 # bal.tab(COVS, treat = test.data$cohort, weights = test.data$weights,
31 v.threshold = 2)
32
33
34
35 # But plotting is clearer:
36 love.plot(COVS, treat = test.data$cohort, weights = test.data$weights,
37 stats = c("mean.diffs"), thresholds = c(m = .1), var.order =
38 "adjusted")
39
40 # We can also plot variance ratios for continuous variables
41 love.plot(COVS, treat = test.data$cohort, weights = test.data$weights,
42 stats = c("variance.ratios"))
43
44 # propensity scores enhance the balance overall, except for the CDAI0.
45 However, this is the reason we use the AIPTW. The remaining imbalance
46 is accounted for by the outcome model (outcome model is the cox
47 regression), and the misspecification of the outcome model is
48 mitigated by the balancing done by propensity score.
49
50
51 First plot to get the difference in average treatment effect in percentage
52
53 plot.ate.diff <- ggplot(dt.out[type == "meanRisk"], aes(x = time,
54 group = level))+
55   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
56 0.4)+
57
58
59
```

```

1
2
3     geom_line(aes(y = estimate, color = level), size = 1)+
4     #geom_vline(xintercept = 90)+
5
6     scale_colour_manual(values = c("lightblue","darkseagreen3"))+
7     scale_fill_manual(values = c("lightblue","darkseagreen3"))+
8     theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
9     legend.position="top" )+
10    scale_x_continuous(breaks=seq(0,500,50)) + scale_y_continuous(labels
11    = scales::percent)+
12
13
14    xlab("Days since intiation of treatment")+
15    ylab("Absolute Risk of treatment discontinuation (%)")+
16    labs(colour="Groups:", fill = "Groups:")+
17    labs(group = "Groups:")+
18    theme_bw(base_size = 14)+
19    theme(axis.title.x = element_text(margin = margin(t = .3,unit =
20    "cm")),
21          axis.title.y = element_text(margin = margin(r = .3,unit =
22    "cm")))
23
24 plot.ate.diff

```

Second plot to get the ratio in average treatment effect

```

28 plot.ate.ratio <- ggplot(dt.out[type == "ratioRisk"], aes(x = time,
29 group = level))+
30   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
31   0.4)+
32   geom_line(aes(y = estimate, color = level), size = 2)+
33
34   theme_minimal()+
35   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
36   scale_x_continuous(breaks=seq(0,500,50))+
37   scale_y_continuous(limits = c(0.9,4.5))+
38
39
40   xlab("Days since intiation of treatment")+
41   ylab("Ratio in Average Treatment Effect")+
42   labs(colour="treatment", fill = "treatment")
43
44 plot.ate.ratio

```

We can also consider the AIPTW estimate at a specific time point. For example at 365-day.

```

48 r.one <- dt.out[type == "diffRisk" & time == 365, .
49 (estimate,lower,upper,p.value)]
50 r.two <- dt.out[type == "ratioRisk" & time == 365, .
51 (estimate,lower,upper,p.value)]
52

```

```

54 ploufrows <- c("Difference in average treatment effect","Ratio in
55 average treatment effect")
56 ploufcols <- c("Estimate","95%CI","p")
57
58
59

```

```

1
2
3 table <- matrix(data = NA, nrow = length(ploufrows), ncol =
4 length(ploufcols))
5 rownames(table) <- ploufrows
6 colnames(table) <- ploufcols
7
8
9 library(formattable)
10 table[1,"Estimate"] <- paste0(formattable(r.one$estimate*100),"%")
11 table[1,"95%CI"] <-
12 paste0(formattable(r.one$lower),"-",formattable(r.one$upper))
13 table[1,"p"] <- writepvalue(r.one$p.value)
14 table[2,"Estimate"] <- paste0(r.two$estimate)
15 table[2,"95%CI"] <- paste0(r.two$lower,"-",r.two$upper)
16 table[2,"p"] <- writepvalue(r.two$p.value)
17
18 table
19
20 # Interpretation: If every patient had received BARI, the 365-day risk
21 of treatment discontinuation would have been 19.34% (points) lower
22 compared to when every patient had received TNFi.

```

BARI vs OMA --

COX model

```

28 BARI2.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
29 cohort+
30
31                               I(age_base/10)+
32                               bmi_base+
33                               TC_with_csDMARD+
34                               PREDNISON_STEROID+
35                               I(CDAI0/10)+
36                               I(disease_duration_base_years/10)+
37                               C(smoker_base, base=3)+
38                               line_of_therapy+
39                               gender+
40                               seropositivity_base+
41                               cluster(patient_id),
42                               fitter = coxph, xtrans = imputed_data2,
43                               data = BARI2)
44
45 summary(BARI2.adj.mi)

```

Creation of HR table and p-values (denis)

```

48 ploufrows <- names(BARI2.adj.mi$coefficients)
49 ploufcols <- c("HR","95%CI","p")
50 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
51 length(ploufcols))
52 rownames(coxtable) <- ploufrows
53 colnames(coxtable) <- ploufcols
54 plouf <- summary(BARI2.adj.mi)

```

```

1
2
3   for(row in ploufrows)
4   {
5     coxtable[row,"HR"] <-
6     formattable(plouf$coefficients[row,"exp(coef)"])
7     coxtable[row,"95%CI"] <-
8     paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",formattable(pl
9     ouf$conf.int[row,"upper .95"]))
10    coxtable[row,"p"] <- writepvalue(plouf$coefficients[row,"Pr(>|
11    z|)"])
12  }
13
14
15  write.xlsx(coxtable, file="./3_clean_output/BARI vs OMA HR.xlsx") #
16  saving excel file
17

```

Forest plot

```

18
19  meanall <- summary(BARI2.adj.mi)$coefficients[1:14,"exp(coef)"]
20  lowerall <- summary(BARI2.adj.mi)$conf.int[1:14,"lower .95"]
21  upperall <- summary(BARI2.adj.mi)$conf.int[1:14,"upper .95"]
22  textall <- c("OMA (vs BARI)", "Age (decades)", "BMI", "Concomitant
23  csDMARD", "Concomitant glucocorticoid", "CDAI score (10 pts)",
24  "Disease duration (decades)", "Current smoker (vs non-smoker)",
25  "Former smoker (vs non-smoker)", "2nd line therapy (vs
26  1st)", "3rd line therapy (vs 1st)", "4th or later line (vs 1st)",
27  "Female gender", "Seropositivity (RF or ACPA)")
28
29
30  dfall <- data.frame(textall, meanall, lowerall, upperall)
31  dfall$textall <- factor(dfall$textall,
32  levels = textall)
33
34  HR_plot_2 <- ggplot(data=dfall, aes(x=textall, y= meanall, ymin =
35  lowerall, ymax = upperall))+
36
37    geom_pointrange(size=0.5)+
38    geom_errorbar(aes(ymin=lowerall, ymax=upperall),width=0.5)+
39    geom_hline(yintercept =1, linetype=2)+
40
41    xlab('')+ ylab(" ")
42    ggtitle("BARI vs OMA")+
43
44    scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
45  +
46  facet_wrap(~textall,nrow=16, strip.position= "right", scales =
47  "free_y") +
48
49
50  theme_pubclean()+
51  theme(strip.text.y = element_blank(),
52  strip.background = element_blank(),
53  axis.line.x = element_line(size = 0.5),
54  axis.text = element_text(face = "bold", colour = "black"),
55  legend.position="bottom", plot.margin =
56
57
58
59

```

```
1
2
3 unit(c(1,3,2,1),"lines"))+
4
5     coord_flip()
6
7 HR_plot_2
8
9
10 # adding some manual annotation
11 grid.text("Improves drug maintenance", x = unit(0.3, "npc"), y =
12 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
13 grid.text("Reduces drug maintenance", x = unit(0.87, "npc"), y =
14 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
15
```

Non-adjusted Kaplan-Meier curves

based on mini-tutorial found on datacamp.com/community/tutorials/survival-analysis-R)

BARI vs OMA

```
21 surv_object2 <- Surv(time = BARI2$time_on_drug, event =
22 BARI2$stop_DMARD)
23 fit2 <- survfit(surv_object2 ~ cohort, data = BARI2) # this function
24 creates the data for Kaplan Meyer
25 survplot_2 <- ggsurvplot(fit2, data = BARI2, # plot
26   pval = T,
27   pval.method = TRUE,
28   legend.title = "Groups :",
29   legend.labs = c("Baricitinib", "OMA"),
30   xlab = "Time (days)",
31   xlim = c(0, 700),
32   censor = FALSE,
33   title = "Non-adjusted model of drug discontinuation by type
34 of treatment",
35   surv.median.line = "v",
36   linetype = 1,
37   size = 1.5,
38   ggtheme = theme_minimal(),
39   #palette = c("grey78", "grey50"),
40   palette = c("red2", "blue3"), # to put colors
41   risk.table = T)
42
43 survplot_2
44 summary(fit2, times = 365)
45 summary(fit2, times = 730)
46
```

Saving the plot curv object for Lilly

```
47
48
49 plot_BARI_vs_OMA_data <- survplot_2$data.survplot
50 write.xlsx(plot_BARI_vs_OMA_data, file =
51 "./3_clean_output/Lilly_curves_excel/plot_BARI_vs_OMA_data_non_adjuste
52 d.xlsx", row.names = F)
53
54
55
56
57
58
59
```

Home-made attempt to obtain adjusted cuves based on imputed data :

```

1
2
3
4 dummy_cox_impute2 <- mice::complete(imputed_data2, "long", include =
5 T)
6 dummy_cox_impute2 <- dummy_cox_impute2[dummy_cox_impute2$.imp != 0,]
7
8
9 BARI_fit2 <- survfit(coxph(Surv(time = time_on_drug, event =
10 stop_DMARD) ~ cohort+
11           I(age_base/10)+
12           bmi_base+
13           TC_with_csDMARD+
14           PREDNISON_STEROID+
15           CDAI0+
16           I(disease_duration_base_years/10)+
17           C(smoker_base, base=3)+
18           line_of_therapy+
19           gender+
20           seropositivity_base+
21           cluster(patient_id)+
22           strata(cohort),dummy_cox_impute2), data =
23 dummy_cox_impute2)
24
25
26 survplot_2_adj <- ggsurvplot(BARI_fit2, data = dummy_cox_impute2,
27 variable = "cohort",
28       xlab = "Time (days)",
29       title = "Multivariable Cox model of drug discontinuation by
30 type of treatment - BARI vs OMA",
31       legend.title = "Groups :",
32       legend.labs = c("Baricitinib", "OMA bDMARDs"),
33       censor = FALSE,
34       xlim = c(0, 700),
35       surv.median.line = "v",
36       linetype = 1,
37       size = 1.5,
38       ggtheme = theme_minimal(),
39       # palette = c("grey78", "grey50")
40       palette = c("red2", "blue3") # to change colors
41     )
42
43
44 # adding some legends
45 survplot_2_adj <- survplot_2_adj +
46   labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
47 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
48 smoking status, line of therapy, gender, seropositivity")
49
50
51 survplot_2_adj
52 summary(BARI_fit2, times = 365) # to see detailed surv probabilities
53 at given timepoints
54 summary(BARI_fit2, times = 730)
55
56 Saving the plot curv object for Lilly
57
58
59
60

```

```

1
2
3 plot_BARI_vs_OMA_data_adj <- survplot_2_adj$data.survplot
4 write.xlsx(plot_BARI_vs_OMA_data_adj, file =
5 ".\3_clean_output\Lilly_curves_excel\plot_BARI_vs_OMA_data_adj.xlsx",
6 row.names = F)
7

```

Sensitivity analysis with package RiskRegression (AIPTW)

```

8
9 # I select only one imputed dataset. Would be good to find a way to
10 pool the results from the 50 datasets imputed.
11

```

```

12 test.data2 <- complete(imputed_data2,1)
13

```

```

14 # First, we specify the treatment model (propensity score model)
15 # Logistic regression where the treatment group is the dependent
16 variable.
17

```

```

18 m.treatment2 <- glm(cohort~I(age_base/10)+
19                    bmi_base+
20                    TC_with_csDMARD+
21                    PREDNISON_STEROID+
22                    I(CDAI0/10)+
23                    I(disease_duration_base_years/10)+
24                    C(smoker_base, base=3)+
25                    line_of_therapy+
26                    gender+
27                    seropositivity_base,
28                    data = test.data2, family =
29                    "binomial" )
30

```

```

31
32 # Then we specify both the "event model" and the "censoring model".
33 Both are cox model
34

```

```

35 m.event2 <- coxph(Surv(time_on_drug, stop_DMARD) ~ cohort+
36                  I(age_base/10)+
37                  bmi_base+
38                  TC_with_csDMARD+
39                  PREDNISON_STEROID+
40                  I(CDAI0/10)+
41                  I(disease_duration_base_years/10)+
42                  C(smoker_base, base=3)+
43                  line_of_therapy+
44                  gender+
45                  seropositivity_base,
46                  data = test.data2, x = TRUE, y =
47                  TRUE)
48

```

```

49
50 m.censor2 <- coxph(Surv(time_on_drug,stop_DMARD==0) ~ cohort +
51                  I(age_base/10)+
52                  bmi_base+
53                  TC_with_csDMARD+
54                  PREDNISON_STEROID+
55                  I(CDAI0/10)+
56

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base
, x =TRUE, y = TRUE,
data = test.data2)

# And we measure the average treatment effect using function "ate",
specifying the times at which we want to compute the ATE

out2 <- ate(event = m.event2 ,
            treatment = m.treatment2,
            censor = m.censor2,
            data = test.data2,
            cause = 1,
            estimator = "AIPTW",
            times = seq(from = 0, to = 500, by = 5))

dt.out2 <- as.data.table(out2)

Diagnostics asked by Lily statistician

library(cobalt)

# First, the distribution of propensity scores
test.data2$pscores <- m.treatment2$fitted.values
test.data2 %>% setDT()

pscore_plot2 <- ggplot(test.data2,aes(x = pscores, color = cohort,
fill = cohort)) +
  geom_density(alpha = .47) +
  xlab("Estimated Probability of being assigned BARI") +
  ylab("Density") +
  theme_minimal()+
  theme(axis.ticks.y=element_blank(),
        panel.grid.minor=element_blank(),
        legend.title=element_blank(),
        text = element_text(size = 16),
        axis.title.x =element_text(hjust = 0.2, size = 16))
pscore_plot2
# Good overlap

## Computing the weights
test.data2$weights <- ifelse(test.data2$cohort == "OMA",
1/test.data2$pscores, 1/(1-test.data2$pscores))

# Selecting only our covariates of interest (the ones in the ps model)
COVS_2 <- subset(test.data2, select = c(cohort,age_base,
bmi_base,

```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

TC_with_csDMARD,
PRÉDNISÓN_STEROID,
CDAI0,
disease_duration_base_years,
smoker_base,
line_of_therapy,
gender,
seropositivity_base))

# To get the SMD & variance ratios before/after weighting
# bal.tab(COVS_2, treat = test.data2$cohort, thresholds = 0.1)
# bal.tab(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, thresholds = 0.1)
# bal.tab(COVS_2, treat = test.data2$cohort, v.threshold = 2)
# bal.tab(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, v.threshold = 2)
#

#But plotting it is better:
love.plot(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
var.order = "adjusted")

# We can also plot variance ratios for continuous variables
love.plot(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights,stats = c("variance.ratios"))

# propensity scores enhance the balance overall, except for the CDAI0.
However, this is the reason we use the AIPTW. The remaining imbalance
is accounted for by the outcome model (outcome model is the cox
regression), and the misspecification of the outcome model is
mitigated by the balancing done by propensity score.

First plot to get the difference in average treatment effect in percentage

plot.ate.diff2 <- ggplot(dt.out2[type == "meanRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.3)+
  geom_line(aes(y = estimate, color = level), size = 1)+

  theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
legend.position="top" )+
  scale_x_continuous(breaks=seq(0,500,50)) + scale_y_continuous(labels
= scales::percent)+

  xlab("Days since intiation of treatment")+
  ylab("Absolute Risk of treatment discontinuation (%)")+
  labs(colour="Groups:", fill = "Groups:", title = "Absolute risk of
treatment discontinuation by type of treatment - BARI vs TNFi")+

```

```

1
2
3     labs(group = "Groups:")
4
5 plot.ate.diff2
6
7 Second plot to get the ratio in average treatment effect
8
9 plot.ate.ratio2 <- ggplot(dt.out2[type == "ratioRisk"], aes(x = time,
10 group = level))+
11   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
12 0.3)+
13   geom_line(aes(y = estimate, color = level), size = 1)+
14
15   theme_minimal()+
16   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
17   scale_x_continuous(breaks=seq(100,400,50))+
18   scale_y_continuous(limits = c(0.8,3))+
19
20   xlab("Days since initiation of treatment")+
21   ylab("Ratio in Average Treatment Effect")+
22   labs(colour="treatment", fill = "treatment")
23
24

```

```

25 plot.ate.ratio2
26

```

We can also consider the AIPTW estimate at a specific time point. For example at 365-day.

```

27
28
29 r.one <- dt.out2[type == "diffRisk" & time == 365, .
30 (estimate,lower,upper,p.value)]
31 r.two <- dt.out2[type == "ratioRisk" & time == 365, .
32 (estimate,lower,upper,p.value)]
33
34
35 ploufrows <- c("Difference in average treatment effect","Ratio in
36 average treatment effect")
37 ploufcols <- c("Estimate","95%CI","p")
38 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
39 length(ploufcols))
40 rownames(coxtable) <- ploufrows
41 colnames(coxtable) <- ploufcols
42
43 library(formattable)
44 coxtable[1,"Estimate"] <- paste0(formattable(r.one$estimate*100),"%")
45 coxtable[1,"95%CI"] <-
46 paste0(formattable(r.one$lower),"-",formattable(r.one$upper))
47 coxtable[1,"p"] <- writepvalue(r.one$p.value)
48 coxtable[2,"Estimate"] <- paste0(r.two$estimate)
49 coxtable[2,"95%CI"] <- paste0(r.two$lower,"-",r.two$upper)
50 coxtable[2,"p"] <- writepvalue(r.two$p.value)
51
52

```

```

53 coxtable
54

```

Interpretation: If every patient had received BARI, the 365-day risk

of treatment discontinuation would have been xx% (points) lower compared to when every patient had received TNFi.

.

[3] 1st LINE vs 1st LINE analysis

Common Table 1

Table 1 with NA, to have exact counts and proportions

```

BARI_first <- BARI_DATA
BARI_first <- BARI_first[line_of_therapy == "1st"] # selection of TC
first line

myVars2 <- c("gender", "age_base", "disease_duration_base_years",
"CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
"seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
"line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
"PREDNISON_STEROID_dose", "dose", "initiation_year",
"time_on_drug", "HAQ_score_base")

catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
"obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
"time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
"seropositivity_base")

nonnormalVars <- c()

tab1 <- CreateTableOne(vars = myVars2, data = BARI_first, factorVars =
catVars2, strata = "cohort", test = F, includeNA = T)
tablexp <- print(tab1, nonnormal= nonnormalVars, catDigits = 1,
contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)

saving

write.xlsx(tablexp, file = "./3_clean_output/BARI 3 groups first line
table1 NA.xlsx")

```

Table 1 without NA to have adequate p values to interpret

```

BARI_first <- BARI_DATA
BARI_first <- BARI_first[line_of_therapy == "1st"] # selection TC
first line

myVars2 <- c("gender", "age_base", "disease_duration_base_years",
"CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
"seropositivity_base", "time_on_drug365", "TC_with_csDMARD",

```

```

1
2
3 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
4 "PREDNISON_STEROID_dose", "dose", "initiation_year",
5 "time_on_drug","HAQ_score_base")
6
7
8 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
9 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
10 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
11 "seropositivity_base")
12
13 nonnormalVars <- c()
14
15 tab1 <- CreateTableOne(vars = myVars2, data = BARI_first, factorVars =
16 catVars2, strata = "cohort", test = T, includeNA = F)
17 tablexp <- print(tab1, nonnormal= nonnormalVars, catDigits = 1,
18 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
19

```

Saving

```

20
21
22 write.xlsx(tablexp, file = "./3_clean_output/BARI 3 groups first line
23 table1.xlsx")
24
25 summary(BARI_first[cohort=="BARI", c("TC_id", "patient_id",
26 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
27 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
28 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
29 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
30 "cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
31 to see NA values for all variables
32
33
34 summary(BARI_first[cohort=="TNFi", c("TC_id", "patient_id",
35 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
36 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
37 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
38 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
39 "cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
40 to see NA values for all variables
41
42
43 summary(BARI_first[cohort=="OMA", c("TC_id", "patient_id",
44 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
45 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
46 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
47 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
48 "cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
49 to see NA values for all variables
50

```

Non-adjusted Survival curves

BARI vs TNFi

```

51
52
53
54
55 BARI_first1 <- copy(BARI_first[cohort %in% c("BARI", "TNFi")])
56
57
58
59

```

```
1
2
3 surv_object3 <- Surv(time = BARI_first1$time_on_drug, event =
4 BARI_first1$stop_DMARD) # indiquer stop variable and time_on_drug
5 summary(coxph(surv_object3 ~ cohort, data = BARI_first1))
6 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first1) # fonction
7 which creates Kaplan-meier data
8 survplot_first1 <- ggsurvplot(fit3, data = BARI_first1, # plot
9     pval = T,
10    pval.method = TRUE,
11    legend.title = "Groups :",
12    legend.labs = c("Baricitinib", "TFNi"),
13    xlab = "Time (days)",
14    xlim = c(0, 700),
15    censor = FALSE,
16    title = "Non-adjusted model of drug discontinuation by type
17 of treatment",
18    surv.median.line = "v",
19    linetype = 1,
20    size = 1.5,
21    ggtheme = theme_minimal(),
22    # palette = c("grey78", "grey50", "grey10"),
23    palette = c("red2", "green3"), # to get colors
24    risk.table = T
25    )
26
27
28
29 survplot_first1
30 table(BARI_first1$cohort)
31 summary(fit3)
32
33 rm(surv_object3, fit3)
34
35 BARI vs OMA
36
37 BARI_first2 <- BARI_first[line_of_therapy == "1st" & cohort %in%
38 c("BARI", "OMA")] # selection des TC TNFi
39
40 surv_object3 <- Surv(time = BARI_first2$time_on_drug, event =
41 BARI_first2$stop_DMARD) # indiquer stop variable and time_on_drug
42 summary(coxph(surv_object3 ~ cohort, data = BARI_first2))
43 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first2) # fonction
44 which creates Kaplan-meier data
45 survplot_first2 <- ggsurvplot(fit3, data = BARI_first2, # plot
46     pval = T,
47    pval.method = TRUE,
48    legend.title = "Groups :",
49    legend.labs = c("Baricitinib", "OMA"),
50    xlab = "Time (days)",
51    xlim = c(0, 700),
52    censor = FALSE,
53    title = "Non-adjusted model of drug discontinuation by type
54 of treatment",
55
56
57
58
59
```

```

1
2
3     surv.median.line = "v",
4     linetype = 1,
5     size = 1.5,
6     ggtheme = theme_minimal(),
7     # palette = c("grey78", "grey50", "grey10"),
8     palette = c("red2", "blue3"), # to get colors
9     risk.table = T
10    )
11

```

```

12
13 survplot_first2
14 table(BARI_first2$cohort)
15 summary(fit3)
16

```

```

17 rm(surv_object3, fit3)
18

```

Adjusted survival analyses

BARI vs TNFi

Verification (quick)

```

25 # Test of proportionality of hazards on raw data
26 test_first_ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
27 as.factor(cohort)+

```

```

28     cluster(patient_id),
29     data= BARI_first1)
30 cox.zph(test_first_ph)
31

```

Adjusted Cox-model

```

33 imputed_data1_first <- complete(imputed_data1,"long",include=T) # to
34 put in the long format
35 imputed_data1_first <- filter(imputed_data1_first, line_of_therapy ==
36 "1st") # only keep 1st line imputed TC
37 imputed_data1_first <- as.mids(imputed_data1_first) # put back in
38 previous format, to use fit.mult.impute
39

```

```

41 BARI_first1.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
42 cohort+

```

```

43     I(age_base/10)+
44     bmi_base+
45     concomitant_csDMARD+
46     PREDNISON_STEROID+
47     I(CDAI0/10)+
48     I(disease_duration_base_years/10)+
49     C(smoker_base, base=3)+
50     line_of_therapy+
51     gender+
52     seropositivity_base+
53     cluster(patient_id),
54     fitter = coxph, xtrans =
55

```

```

1
2
3 imputed_data1_first, data = BARI_first1)
4 summary(BARI_first1.adj.mi)
5
6 Creation of HR table with p-values
7 ploufrows <- names(BARI_first1.adj.mi$coefficients)
8 ploufcols <- c("HR","95%CI","p")
9 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
10 length(ploufcols))
11 rownames(coxtable) <- ploufrows
12 colnames(coxtable) <- ploufcols
13 plouf <- summary(BARI_first1.adj.mi)
14
15
16 for(row in ploufrows)
17 {
18   coxtable[row,"HR"] <-
19   formattable(plouf$coefficients[row,"exp(coef)"])
20   coxtable[row,"95%CI"] <-
21   paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",formattable(pl
22 ouf$conf.int[row,"upper .95"]))
23   coxtable[row,"p"] <- writepvalue(plouf$coefficients[row,"Pr(>|
24 z|)"])
25 }
26
27
28 write.xlsx(coxtable, file="./3_clean_output/BARI vs TNFi HR 1st
29 lines.xlsx") # save in excel format
30
31 Adjusted curves with imputed data
32 dummy_cox_impute_first1 <- mice::complete(imputed_data1_first, "long",
33 include = T)
34 dummy_cox_impute_first1 <-
35 dummy_cox_impute_first1[dummy_cox_impute_first1$.imp != 0,]
36
37 BARI_first1_fit <- survfit(coxph(Surv(time = time_on_drug, event =
38 stop_DMARD) ~ cohort+
39
40 I(age_base/10)+
41 bmi_base+
42 concomitant_csDMARD+
43 PREDNISON_STEROID+
44 CDAI0+
45 I(disease_duration_base_years/10)+
46 C(smoker_base, base=3)+
47 line_of_therapy+
48 gender+
49 seropositivity_base+
50 cluster(patient_id)+
51 strata(cohort),dummy_cox_impute_first1),
52 data = dummy_cox_impute_first1)
53
54
55
56 survplot_first1_adj <- ggsurvplot(BARI_first1_fit, data =
57
58
59
60

```

```

1
2
3 dummy_cox_impute_first1, variable = "cohort",
4     xlab = "Time (days)",
5     title = "Multivariable Cox model of drug discontinuation by
6 type of treatment - 1st line vs 1st line",
7     legend.title = "Groups :",
8     legend.labs = c("Baricitinib", "TNFi"),
9     censor = FALSE,
10    xlim = c(0, 700),
11    surv.median.line = "v",
12    linetype = 1,
13    size = 1.5,
14    ggtheme = theme_minimal(),
15    #palette = c("grey78", "grey10")
16    palette = c("red2", "green3"), # to get colors
17    )
18
19
20 survplot_first1_adj <- survplot_first1_adj +
21     labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
22 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
23 smoking status, line of therapy, gender, seropositivity")
24
25

```

```

26 survplot_first1_adj
27 table(BARI_first1$cohort)
28 rm(dummy_cox_impute_first1, BARI_first1_fit)
29

```

BARI vs OMA

Verification (quick)

```

31
32
33 # Test of proportionality of hazards on raw data
34 test_first_ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
35 as.factor(cohort)+
36
37     cluster(patient_id),
38     data= BARI_first2)
39 cox.zph(test_first_ph)
40

```

Adjusted Cox-model

```

41
42 imputed_data2_first <- complete(imputed_data2,"long",include=T) # to
43 put in the long format
44 imputed_data2_first <- filter(imputed_data2_first, line_of_therapy ==
45 "1st") # only keep 1st line imputed TC
46 imputed_data2_first <- as.mids(imputed_data2_first) # put back in
47 previous format, to use fit.mult.impute
48
49
50 BARI_first2.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
51 cohort+
52     I(age_base/10)+
53     bmi_base+
54     concomitant_csDMARD+
55     PREDNISON_STEROID+
56
57
58
59

```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

I(CDAI0/10)+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base+
cluster(patient_id),
fitter = coxph, xtrans =

```

imputed_data2_first, data = BARI_first2)
summary(BARI_first2.adj.mi)

```

Creation of HR table with p-values

```

ploufrows <- names(BARI_first2.adj.mi$coefficients)
ploufcols <- c("HR", "95%CI", "p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols
plouf <- summary(BARI_first2.adj.mi)

for(row in ploufrows)
{
  coxtable[row, "HR"] <-
formattable(plouf$coefficients[row, "exp(coef)"])
  coxtable[row, "95%CI"] <-
paste0(formattable(plouf$conf.int[row, "lower .95"]), "- ", formattable(pl
ouf$conf.int[row, "upper .95"]))
  coxtable[row, "p"] <- writepvalue(plouf$coefficients[row, "Pr(>|
z|)"])
}

write.xlsx(coxtable, file="./3_clean_output/BARI vs OMA HR 1st
lines.xlsx") # save in excel format

```

Adjusted curves with imputed data

```

dummy_cox_impute_first2 <- mice::complete(imputed_data2_first, "long",
include = T)
dummy_cox_impute_first2 <-
dummy_cox_impute_first2[dummy_cox_impute_first2$.imp != 0,]

BARI_first2_fit <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+

```

I(age_base/10)+
bmi_base+
concomitant_csDMARD+
PREDNISON_STEROID+
CDAI0+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+

```

1
2
3           gender+
4           seropositivity_base+
5           cluster(patient_id)+
6           strata(cohort), dummy_cox_impute_first2),
7 data = dummy_cox_impute_first2)
8
9
10
11 survplot_first2_adj <- ggsurvplot(BARI_first2_fit, data =
12 dummy_cox_impute_first2, variable = "cohort",
13   xlab = "Time (days)",
14   title = "Multivariable Cox model of drug discontinuation by
15 type of treatment - 1st line vs 1st line",
16   legend.title = "Groups :",
17   legend.labs = c("Baricitinib", "OMA bDMARDs"),
18   censor = FALSE,
19   xlim = c(0, 700),
20   surv.median.line = "v",
21   linetype = 1,
22   size = 1.5,
23   ggtheme = theme_minimal(),
24   #palette = c("grey78", "grey50")
25   palette = c("red2", "blue3"), # to get colors
26 )
27
28
29 survplot_first2_adj <- survplot_first2_adj +
30   labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
31 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
32 smoking status, line of therapy, gender, seropositivity")
33
34 survplot_first2_adj
35 table(BARI_first2$cohort)
36 rm(dummy_cox_impute_first2, BARI_first2_fit)
37
38
39

```

1. [4] LACK of EFFICACY and ADVERSE EVENTS

Analysis by stop_reasons in competing risk

(BARI vs TNFi)

Cumulative incidence function

```
BARI_comp <- copy(BARI_DATA)
```

#General

```

51
52 BARI_comp[stop_reasons == "ADVERSE_EVENT", status := 1]
53 BARI_comp[stop_reasons == "NOT_EFFECTIVE", status := 2]
54 BARI_comp[stop_reasons == "OTHER" | stop_reasons == "REMISSION",
55 status := 3]
56
57
58
59

```

```

1
2
3 BARI_comp[stop_reasons == "CONTINUE", status := 0]
4 BARI_comp$cohort <- as.factor(BARI_comp$cohort)
5
6
7 library(reshape)
8
9 BARI_comp_B <- BARI_comp[cohort %in% c("BARI")] #BARI only
10 ci_BARI <- Cuminc(time = "time_on_drug", status = "status", data =
11 BARI_comp_B)
12 ci_BARI <- ci_BARI[, -c(2,6,7,8,9)]
13 ci_long_BARI <- reshape2::melt(ci_BARI, id.vars = "time")
14
15 BARI_comp_T <- BARI_comp[cohort %in% c("TNFi")] #TNFi only
16 ci_TNFi <- Cuminc(time = "time_on_drug", status = "status", data =
17 BARI_comp_T)
18 ci_TNFi <- ci_TNFi[, -c(2,6,7,8,9)]
19 ci_long_TNFi <- reshape2::melt(ci_TNFi, id.vars = "time")
20
21
22 ci_long_BARI$cohort <- 0
23 ci_long_TNFi$cohort <- 1
24 ci_long <- rbind(ci_long_BARI, ci_long_TNFi)
25 ci_long$cohort <- as.factor(ci_long$cohort)
26
27 plot2 <- ggplot(data = ci_long, aes(x = time,
28 y = value,
29 linetype =
30 interaction(cohort, variable),
31 col =
32 interaction(cohort, variable))) +
33 geom_line(size = 0.75) +
34 scale_color_manual(name = "",
35 values
36 =c("#08306B", "#08306B", "#238B45", "#238B45", "#FD8D3C", "#FD8D3C"),
37 labels = c("Adverse Event (BARI)", "Adverse Event
38 (TNFi)", "Ineffectiveness (BARI)", "Ineffectiveness (TNFi)", "Other
39 (BARI)", "Other (TNFi)"))+
40 scale_linetype_manual(name="",
41 values = c(1,3,1,3,1,3),
42 labels = c("Adverse Event (BARI)", "Adverse
43 Event (TNFi)", "Ineffectiveness (BARI)", "Ineffectiveness (TNFi)", "Other
44 (BARI)", "Other (TNFi)"))+
45 scale_x_continuous(name = "Time", limits = c(1,365)) +
46 scale_y_continuous(name = "Cumulative incidence", limits =
47 c(0.0,0.3)) +
48 theme_bw()+
49 theme(strip.text.y = element_blank(),
50 strip.background = element_blank(),
51 axis.line.x = element_line(size = 0.5),
52 axis.text = element_text(face = "bold", colour = "black"),
53 legend.position="right", plot.margin =
54 unit(c(1,3,2,1),"lines"))+
55
56
57
58
59
60

```

```

1
2
3     #ggtitle("Cumulative incidence functions")+
4     theme(plot.title = element_text(hjust = 0.5))
5
6 plot2
7
8 Adjusting variables
9 # Covariates of interest for Cox
10
11 covs <-
12 c("cohort","age_base","bmi_base","TC_with_csDMARD","PREDNISON_STEROID"
13   ,"CDAI0","disease_duration_base_years","smoker_base","line_of_therapy"
14   ,"gender","seropositivity_base")
15
16 Cause-specific hazard model
17 # Rappel: imputed_data1 = BARI vs TNFi
18 #         imputed_data2 = BARI vs OMA
19
20
21 # Transition matrix definition
22 tmat <- trans.comprisk(2, names = c("event-free","ae","lae"))
23 tmat
24
25
26 imputed_data1_long <- complete(imputed_data1, action = "long") %>%
27 setDT()
28 imputed_data1_long[,stop_ae := fifelse(stop_reasons ==
29   "ADVERSE_EVENT",1,0)]
30 imputed_data1_long[,stop_lae := fifelse(stop_reasons ==
31   "NOT_EFFECTIVE",1,0)]
32 imputed_data1_long[,stop_other := fifelse(stop_reasons == "OTHER" |
33   stop_reasons == "REMISSION",1,0)]
34 #[,continue := fifelse(stop_reasons == "CONTINUE",1,0)]
35 imputed_data1_long[,continue := fifelse(stop_reasons == "OTHER" |
36   stop_reasons == "REMISSION" | stop_reasons == "CONTINUE",1,0)]
37
38 M <- imputed_data1$m
39
40
41 mice_fit <- lapply(1:M,function(m){
42
43   # subset
44   data_sub <- imputed_data1_long[.imp == m]
45
46   mst_hosp <- msprep(time =
47   c("time_on_drug","time_on_drug","time_on_drug"),
48     status = c("continue","stop_ae","stop_lae"),
49     data = as.data.frame(data_sub),
50     trans = tmat,
51     keep = covs)
52
53   # get covariates
54   tmp <- expand.covs(mst_hosp,covs, append = TRUE, longnames = T)
55   tmp_cov <- grep(paste0(covs,".",collapse = "|"),names(tmp),value = T)
56
57
58
59
60

```

```

1
2
3   # fit
4   coxph(as.formula(paste0("Surv(Tstart, Tstop, status) ~",
5                       paste0(tmp_cov,collapse = " + "),
6                       "+ strata(trans)")),
7
8       data = tmp,
9       method = "breslow")
10
11  }) %>%
12    as.mira()
13
14  est <- pool(mice_fit)
15
16  # Transition 1 = Adverse Event. Hazard Ratio vs VARI
17  # Transition 2 = Lack of Efficacy. Hazard Ratio vs VARI
18  # estimate = Hazard ratio
19
20  summary(est, conf.int = T, exponentiate = T)
21
22
23  # Conclusion
24  # => The hazard ratio of lack of efficacy (lae) for TNFi is 65% higher
25  than for BARI. Significant.
26  # => No difference between TNFi and BARI for Adverse Event (ae)
27
28  Clean table with confidence intervals & p-values
29
30  # Hazard ratios
31  ploufrows <- as.character(summary(est)$term)
32
33  ploufcols <- c("HR","95%CI","p")
34  coxtable_csh <- matrix(data = NA, nrow = length(ploufrows), ncol =
35  length(ploufcols))
36  rownames(coxtable_csh) <- ploufrows
37  colnames(coxtable_csh) <- ploufcols
38  plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
39
40  for(row in ploufrows)
41  {
42    coxtable_csh[row,"HR"] <- formattable(plouf[term %in% row,
43    estimate])
44    coxtable_csh[row,"95%CI"] <- paste0(formattable(plouf[term %in%
45    row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
46    coxtable_csh[row,"p"] <- writepvalue(plouf[term %in% row,
47    p.value])}
48
49
50  output <- coxtable_csh
51  row.names(output)[1:2] <- c("TNFi Adverse Event (vs BARI)", "TNFi Lack
52  of Eff (vs BARI)")
53
54
55  # Transition 1 = Adverse Event. Hazard Ratio vs VARI
56
57
58
59
60

```

```
1
2
3 # Transition 2 = Lack of Efficacy. Hazard Ratio vs VARI
4 output
5
6 Subdistribution hazard model (Fine-Gray)
7 # Status variable
8 imputed_data1_long[stop_reasons == "ADVERSE_EVENT",status := 1]
9 imputed_data1_long[stop_reasons == "NOT_EFFECTIVE", status := 2]
10 imputed_data1_long[stop_reasons == "OTHER" | stop_reasons ==
11 "REMISSION", status := 3]
12 imputed_data1_long[stop_reasons == "CONTINUE", status := 0]
13
14
15 ## ATTENTION levels() re-ecrit juste l'étiquette!! Change pas la
16 donnée !!! Donc ça re écrit les labels
17
18 imputed_data1_long$line_of_therapy <-
19 as.factor(imputed_data1_long$line_of_therapy)
20 imputed_data1_long$seropositivity_base <-
21 as.factor(imputed_data1_long$seropositivity_base)
22
23
24 levels(imputed_data1_long$cohort) <- c("0","1")
25 levels(imputed_data1_long$line_of_therapy) <- c("0","1","2","3")
26 levels(imputed_data1_long$gender) <- c("0","1")
27 levels(imputed_data1_long$smoker_base) <- c("2","1","0")
28 levels(imputed_data1_long$smoker_base)
29 levels(imputed_data1_long$seropositivity_base) <- c("0","1")
30
31 M <- imputed_data1$m
32
33 # First loop to get estimates for event = 1: ADVERSE EVENT
34
35 mice_fit <- lapply(1:M,function(m){
36   # subset
37   BARI_toto <- imputed_data1_long[.imp == m]
38   # subdistribution hazard model
39   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
40 BARI_toto[,..covs],failcode = 1,cencode = 0)
41
42
43 }) %>%
44   as.mira()
45 est <- pool(mice_fit)
46 summary(est, conf.int = T, exponentiate = T)
47
48 # Second loop to get estimates for event = 2: LACK OF EFFICACY
49
50 mice_fit2 <- lapply(1:M,function(m){
51   # subset
52   BARI_toto <- imputed_data1_long[.imp == m]
53   #subdistribution hazard model
54   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
```

```
1
2
3 BARI_toto[,..covs],failcode = 2,cencode = 0)
4
5 }) %>%
6   as.mira()
7
8
9 est2 <- pool(mice_fit2)
10 summary(est2, conf.int = T, exponentiate = T)
11
12 # Conclusions:
13 # No significant difference in incidence of adverse event between TNFi
14 and BARI
15 # Increased incidence of lack of efficacy for TNFi compared to BARI.
16 # CAREFUL: using a Fine-Gray model allows us to make claim about the
17 association between a covariate and the direction of the increase in
18 incidence, but we can't quantify the magnitude of the increase in
19 incidence.
20
21 Clean tables with Hazard ratios with confidence intervals & p-values
22
23 # Adverse event
24 ploufrows <- as.character(summary(est)$term)
25 ploufcols <- c("HR","95%CI","p")
26 coxtable_ae <- matrix(data = NA, nrow = length(ploufrows), ncol =
27 length(ploufcols))
28 rownames(coxtable_ae) <- ploufrows
29 colnames(coxtable_ae) <- ploufcols
30 plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
31
32
33 for(row in ploufrows)
34 {
35   coxtable_ae[row,"HR"] <- formattable(plouf[term %in% row,
36 estimate])
37   coxtable_ae[row,"95%CI"] <- paste0(formattable(plouf[term %in% row,
38 `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
39   coxtable_ae[row,"p"] <- writepvalue(plouf[term %in% row, p.value])}
40
41 row.names(coxtable_ae)[1] <- c("TNFi vs BARI Advserse Events")
42
43
44 # Lack of efficacy
45 ploufrows <- as.character(summary(est2)$term)
46 ploufcols <- c("HR","95%CI","p")
47 coxtable_lae <- matrix(data = NA, nrow = length(ploufrows), ncol =
48 length(ploufcols))
49 rownames(coxtable_lae) <- ploufrows
50 colnames(coxtable_lae) <- ploufcols
51 plouf <- summary(est2, conf.int = T, exponentiate = T) %>% setDT()
52
53
54 for(row in ploufrows)
55 {
56   coxtable_lae[row,"HR"] <- formattable(plouf[term %in% row,
```

```

1
2
3 estimate])
4   coxtable_lae[row,"95%CI"] <- paste0(formatable(plouf[term %in%
5 row, `2.5 %`]),"-",formatable(plouf[term %in% row, `97.5 %`]))
6   coxtable_lae[row,"p"] <- writepvalue(plouf[term %in% row,
7 p.value])}
8
9
10 row.names(coxtable_lae)[1] <- c("TNFi vs BARI Lack of Eff")
11
12 # output
13 coxtable_ae
14 coxtable_lae
15
16 write.xlsx(coxtable_ae, file="./3_clean_output/BARI vs TNFi HR
17 competing risk Fine-Gray AE.xlsx") # saving excel file
18 write.xlsx(coxtable_lae, file="./3_clean_output/BARI vs TNFi HR
19 competing risk Fine-Gray LAE.xlsx") # saving excel file
20
21 (BARI vs OMA)
22
23 Cumulative incidence function
24 BARI_comp <- copy(BARI_DATA)
25
26 #General
27
28
29 BARI_comp[stop_reasons == "ADVERSE_EVENT",status := 1]
30 BARI_comp[stop_reasons == "NOT_EFFECTIVE", status := 2]
31 BARI_comp[stop_reasons == "OTHER" | stop_reasons == "REMISSION",
32 status := 3]
33 BARI_comp[stop_reasons == "CONTINUE", status := 0]
34 BARI_comp$cohort <- as.factor(BARI_comp$cohort)
35
36
37 library(reshape)
38
39 BARI_comp_B <- BARI_comp[cohort %in% c("BARI")] #BARI only
40 ci_BARI <- Cuminc(time = "time_on_drug",status = "status", data =
41 BARI_comp_B)
42 ci_BARI <- ci_BARI[, -c(2,6,7,8,9)]
43 ci_long_BARI <- reshape2::melt(ci_BARI,id.vars = "time")
44
45 BARI_comp_O <- BARI_comp[cohort %in% c("OMA")] #OMA only
46 ci_OMA <- Cuminc(time = "time_on_drug",status = "status", data =
47 BARI_comp_O)
48 ci_OMA <- ci_OMA[, -c(2,6,7,8,9)]
49 ci_long_OMA <- reshape2::melt(ci_OMA,id.vars = "time")
50
51
52 ci_long_BARI$cohort <- 0
53 ci_long_OMA$cohort <- 1
54 ci_long_2 <- rbind(ci_long_BARI,ci_long_OMA)
55 ci_long_2$cohort <- as.factor(ci_long_2$cohort)
56
57
58
59
60

```



```

1
2
3
4 plot3 <- ggplot(data = ci_long_2, aes(x = time,
5                                     y = value,
6                                     linetype =
7                                     interaction(cohort,variable),
8                                     col =
9                                     interaction(cohort,variable))) +
10   geom_line(size = 0.75) +
11   scale_color_manual(name = "",
12                     values =
13   c("#08306B", "#08306B", "#238B45", "#238B45", "#FD8D3C", "#FD8D3C"),
14   labels = c("Adverse Event (BARI)", "Adverse Event
15   (OMA)", "Ineffectiveness (BARI)", "Ineffectiveness (OMA)", "Other
16   (BARI)", "Other (OMA)"))+
17   scale_linetype_manual(name="",
18   values = c(1,3,1,3,1,3),
19   labels = c("Adverse Event (BARI)", "Adverse
20   Event (OMA)", "Ineffectiveness (BARI)", "Ineffectiveness (OMA)", "Other
21   (BARI)", "Other (OMA)"))+
22   scale_x_continuous(name = "Time", limits = c(1,365)) +
23   scale_y_continuous(name = "Cumulative incidence", limits =
24   c(0.0,0.3)) +
25   theme_bw()+
26   theme(strip.text.y = element_blank(),
27   strip.background = element_blank(),
28   axis.line.x = element_line(size = 0.5),
29   axis.text = element_text(face = "bold", colour = "black"),
30   legend.position="right", plot.margin =
31   unit(c(1,3,2,1),"lines"))+
32   #ggtitle("Cumulative incidence functions")+
33   theme(plot.title = element_text(hjust = 0.5))
34
35
36

```

```

37 plot3
38

```

Adjusting variables

```

39 # Covariates of interest for Cox
40

```

```

41
42 covs <-
43 c("cohort", "age_base", "bmi_base", "TC_with_csDMARD", "PREDNISON_STEROID"
44   , "CDAI0", "disease_duration_base_years", "smoker_base", "line_of_therapy"
45   , "gender", "seropositivity_base")
46

```

Cause-specific hazard model

```

47
48 # Rappel: imputed_data1 = BARI vs TNFi
49 #           imputed_data2 = BARI vs OMA
50

```

```

51 # Transition matrix definition
52

```

```

53 library(mstate)
54 tmat <- trans.comprisk(2, names = c("event-free", "ae", "lae"))
55 tmat
56
57
58
59

```

```

1
2
3
4 imputed_data2_long <- complete(imputed_data2, action = "long") %>%
5 setDT()
6 imputed_data2_long[,stop_ae := fifelse(stop_reasons ==
7 "ADVERSE_EVENT",1,0)]
8 imputed_data2_long[,stop_lae := fifelse(stop_reasons ==
9 "NOT_EFFECTIVE",1,0)]
10
11 imputed_data2_long[,stop_other := fifelse(stop_reasons == "OTHER" |
12 stop_reasons == "REMISSION",1,0)]
13 #[,continue := fifelse(stop_reasons == "CONTINUE",1,0)]
14 imputed_data2_long[,continue := fifelse(stop_reasons == "OTHER" |
15 stop_reasons == "REMISSION" | stop_reasons == "CONTINUE",1,0)]
16
17 M <- imputed_data2$m
18
19 mice_fit <- lapply(1:M,function(m){
20
21   # subset
22   data_sub <- imputed_data2_long[.imp == m]
23
24   mst_hosp <- msprep(time =
25 c("time_on_drug","time_on_drug","time_on_drug"),
26                       status = c("continue","stop_ae","stop_lae"),
27                       data = as.data.frame(data_sub),
28                       trans = tmat,
29                       keep = covs)
30
31   # get covariates
32 tmp <- expand.covs(mst_hosp,covs, append = TRUE, longnames = T)
33 tmp_cov <- grep(paste0(covs,".",collapse = "|"),names(tmp),value = T)
34
35   # fit
36 coxph(as.formula(paste0("Surv(Tstart, Tstop, status) ~",
37                         paste0(tmp_cov,collapse = " + "),
38                         "+ strata(trans)")),
39       data = tmp,
40       method = "breslow")
41
42
43 }) %>%
44   as.mira()
45
46 est <- pool(mice_fit)
47 summary(est, conf.int = T, exponentiate = T)
48
49 # Transition 1 = Adverse Event
50 # Transition 2 = Lack of Efficacy
51
52
53 # Conclusion
54 # => No difference between OMA and BARI for Adverse Event (ae) and for
55 Lack of Event (lae)
56
57
58
59
60

```

Cleaner table with Hazard ratios with confidence intervals & p-values

```

1 ploufrows <- as.character(summary(est)$term)
2 ploufcols <- c("HR", "95%CI", "p")
3 coxtable_csh2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
4 length(ploufcols))
5 rownames(coxtable_csh2) <- ploufrows
6 colnames(coxtable_csh2) <- ploufcols
7 plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
8
9 for(row in ploufrows)
10 {
11   coxtable_csh2[row, "HR"] <- formattable(plouf[term %in% row,
12 estimate])
13   coxtable_csh2[row, "95%CI"] <- paste0(formattable(plouf[term %in%
14 row, `2.5 %`]), "-", formattable(plouf[term %in% row, `97.5 %`]))
15   coxtable_csh2[row, "p"] <- writepvalue(plouf[term %in% row,
16 p.value])}
17
18 row.names(coxtable_csh2)[1:2] <- c("OMA vs BARI Adverse event", "OMA
19 vs BARI Lack of Eff" )
20 coxtable_csh2

```

Subdistribution hazard model (Fine-Gray)

```

21 # Status variable
22 imputed_data2_long[stop_reasons == "ADVERSE_EVENT", status := 1]
23 imputed_data2_long[stop_reasons == "NOT_EFFECTIVE", status := 2]
24 imputed_data2_long[stop_reasons == "OTHER" | stop_reasons ==
25 "REMISSION", status := 3]
26 imputed_data2_long[stop_reasons == "CONTINUE", status := 0]
27
28 imputed_data2_long$line_of_therapy <-
29 as.factor(imputed_data2_long$line_of_therapy)
30 imputed_data2_long$seropositivity_base <-
31 as.factor(imputed_data2_long$seropositivity_base)
32 levels(imputed_data2_long$cohort) <- c("0", "1")
33 levels(imputed_data2_long$line_of_therapy) <- c("0", "1", "2", "3")
34 levels(imputed_data2_long$gender) <- c("0", "1")
35 levels(imputed_data2_long$smoker_base) <- c("2", "1", "0")
36 levels(imputed_data2_long$smoker_base)
37 levels(imputed_data2_long$seropositivity_base) <- c("0", "1")

```

```

38 M <- imputed_data2$m

```

```

39 # First loop to get estimates for event = 1: ADVERSE EVENT

```

```

40 mice_fit <- lapply(1:M, function(m){
41   # subset
42   BARI_toto <- imputed_data2_long[.imp == m]

```

```

1
2
3     #subdistribution hazard model
4     shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
5     BARI_toto[,..covs],failcode = 1,cencode = 0)
6
7   }) %>%
8     as.mira()
9   est <- pool(mice_fit)
10  summary(est, conf.int = T, exponentiate = T)
11
12
13 # Second loop to get estimates for event = 2: LACK OF EFFICACY
14 mice_fit2 <- lapply(1:M,function(m){
15
16   # subset
17   BARI_toto <- imputed_data2_long[.imp == m]
18
19   # subdistribution hazard model
20   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
21   BARI_toto[,..covs],failcode = 2,cencode = 0)
22
23 }) %>%
24   as.mira()
25 est2 <- pool(mice_fit2)
26 summary(est2, conf.int = T, exponentiate = T)
27
28
29 # Conclusions:
30 # No significant difference in incidence of "adverse event" and "lack
31 of efficacy" between TNFi and BARI
32
33 # CAREFUL: using a Fine-Gray model allows us to make claim about the
34 association between a covariate and the direction of the increase in
35 incidence, but we can't quantify the magnitude of the increase in
36 incidence.
37
38 Cleaner Table with Hazard ratios with confidence intervals & p-values
39
40 # Adverse event
41 ploufrows <- as.character(summary(est)$term)
42 ploufcols <- c("HR","95%CI","p")
43 coxtable_ae2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
44 length(ploufcols))
45 rownames(coxtable_ae2) <- ploufrows
46 colnames(coxtable_ae2) <- ploufcols
47 plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
48
49
50 for(row in ploufrows)
51 {
52   coxtable_ae2[row,"HR"] <- formattable(plouf[term %in% row,
53 estimate])
54   coxtable_ae2[row,"95%CI"] <- paste0(formattable(plouf[term %in%
55 row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
56
57
58
59
60

```

```
1
2
3     coxtable_ae2[row,"p"] <- writepvalue(plouf[term %in% row,
4 p.value])}
5
6 row.names(coxtable_ae2)[1] <- c("OMA vs BARI Advserere Events")
7
8
9 # Lack of efficacy
10 ploufrows <- as.character(summary(est2)$term)
11 ploufcols <- c("HR","95%CI","p")
12 coxtable_lae2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
13 length(ploufcols))
14 rownames(coxtable_lae2) <- ploufrows
15 colnames(coxtable_lae2) <- ploufcols
16 plouf <- summary(est2, conf.int = T, exponentiate = T) %>% setDT()
17
18 for(row in ploufrows)
19 {
20   coxtable_lae2[row,"HR"] <- formattable(plouf[term %in% row,
21 estimate])
22   coxtable_lae2[row,"95%CI"] <- paste0(formattable(plouf[term %in%
23 row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
24   coxtable_lae2[row,"p"] <- writepvalue(plouf[term %in% row,
25 p.value])}
26
27
28 row.names(coxtable_lae2)[1] <- c("OMA vs BARI Lack of Eff")
29
30 #Output
31 coxtable_ae2
32 coxtable_lae2
33
34
35 write.xlsx(coxtable_ae2, file="./3_clean_output/BARI vs OMA HR
36 competing risk Fine-Gray AE.xlsx") # saving excel file
37 write.xlsx(coxtable_lae2, file="./3_clean_output/BARI vs OMA HR
38 competing risk Fine-Gray LAE.xlsx") # saving excel file
39
40
```

1. Saving

```
41 save.image(file="./3_clean_output/full_workspaces/workspace_1.RData")
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
```

2 - LDA and REM ANALYSIS

10/11/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)

rm(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))

load("../1_datamanaged_files/datamanaged.Rdata")
```

This code aims at providing estimates for the remission rates of the different treatments groups REM = REMission LDA = Low Disease Activity

Both outcome are base on the CDAI CDAI = Clinical Disease Activity Index

CDAI is an index computed by the physician, which scores the severity of the disease.

1. [0] Exploration

See all available raw CDAI measures :

```
BARI_long[, group := "non-BARI"]
BARI_long[drug == "BIOLOGIC_BARICITINIB", group := "BARI"]

nrow(BARI_DATA)
summary(BARI_DATA[, .(CDAI0_raw, CDAI12_raw)])
```

1. [1] CARRAC (confirm covariates for confounding and for attrition)

For LDA with updated function

```
library(modules)
source_comp_eff <- modules::use("ETAPE_2_supp_code.R")

LDA_BARI_TNF <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "TNFi")],
  var = "CDAI12",
  thres = 10,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons" ),
  seed = 123)

LDA_BARI_OMA <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "OMA")],
  var = "CDAI12",
  thres = 10,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons"),
  seed = 123)

LDA_BARI_TNF
LDA_BARI_OMA
```

For REM with updated function

```
REM_BARI_TNF <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "TNFi")],
  var = "CDAI12",
  thres = 2.8,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons" ),
  seed = 123)

REM_BARI_OMA <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "OMA")],
```

```
1
2
3   var = "CDAI12",
4   thres = 2.8,
5   ttt_var = "cohort",
6   ref_ttt = "BARI",
7   counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
8                   "line_of_therapy", "CDAI0"),
9   attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
10               "line_of_therapy", "CDAI0", "stop_reasons"),
11   seed = 123)
```

```
13
14 REM_BARI_TNF
15 REM_BARI_OMA
```

16
17 This methods was developed by Mongin et al,
18 <https://ard.bmj.com/content/early/2022/01/12/annrheumdis-2021-221477>

20 Pooled table

```
21 table <- rbind(LDA_BARI_TNF, LDA_BARI_OMA, REM_BARI_TNF, REM_BARI_OMA)
22
23 write.xlsx(table, file = "./3_clean_output/table_LDA_REM_CARRAC.xlsx",
24            row.names = F)
```

28 1. Saving

```
29 save.image(file="./3_clean_output/full_workspaces/workspace_2.RData")
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```


2 - LDA and REM supp CODE

10/11/2020

```
{r setup, include=FALSE} import("data.table") import("plyr")
import("data.table") import("mice") import("ipw") import("survey")
import("geepack") import("futile.logger") import("emmeans")
import("stats") import("survival")
```

function to perform checks on data

```
``{r setup, include=FALSE} check_data = function(datain, var = "CDAI_fu", ttt_var = "ttt",
ref_ttt = "ttt_ref", ID_ttt = NULL, othervar = c())
{
  data <- setDT(copy(datain))
  vartochek <- Reduce(union,list(var,ttt_var,othervar)) notindata <-
  setdiff(vartochek,names(data))
  if(length(notindata)>0){ stop(paste0("the variables",paste0(notindata,collapse = ",")," are
  not in the dataplease correct")) }
  # force ttt as var name setnames(data,ttt_var,"ttt")
  if( data[,uniqueN(ttt)]>2){ stop("there are more than two treatments. The analysis has
  been implemented only for 2 treatments") }
  if(!any(data$ttt == ref_ttt)){ stop(paste0("The variable",ttt_var," does not contain any
  ",ttt_ref," value")) }
  data[,ttt := relevel(as.factor(ttt),ref_ttt)] if(is.null(ID_ttt))
  { data[,ID_ttt := .I] }else{ setnames(data,ID_ttt,"ID_ttt") data[,N := .N,by = ID_ttt]
  if(any(data$N>1)){ stop("there are",data[N>1,uniqueN(ID_ttt)]," treatment course which
  have more than one entry in the table. Each row should be an unique treatment") } }
  return(data) }
adjusted_model = function(data, weights = NULL, covariates = NULL){
  # transform char to factor to_fact <- data[,lapply(.SD,class)] %>% transpose(keep.names =
  "var") %>% .[V1 == "character",var]
  data[,c(to_fact) := lapply(.SD,factor),.SDcols = to_fact]
  #droplevels facto_vars <- data[,lapply(.SD,class)] %>% transpose(keep.names = "var") %>
  % .[V1 == "factor",var] data[,c(facto_vars) := lapply(.SD,droplevels),.SDcols = facto_vars]
```

```

1
2
3 # define formula formula <- as.formula(paste0("LDA ~",paste0(c("ttt",covariates),collapse
4 = " + "))
5
6 if(!is.null(covariates)){ # fit fit <- geeglm(formula, data = data, id = ID_ttt, family =
7 gaussian) }else{ fit <- geeglm(LDA ~ ttt, data = data, weights = weights, id = ID_ttt, family =
8 gaussian) }
9
10 fitsummary <- summary(fit) # create table with difference between the two treatments diff
11 <- data.table(ttt = "diff", LDA = fitsummary$coefficients[2,"Estimate"], LDA_var =
12 fitsummary$coefficients[2,"Std.err"]^2, LDA_sup = fitsummary$coefficients[2,"Estimate"] +
13 1.96*fitsummary$coefficients[2,"Std.err"], LDA_inf = fitsummary$coefficients[2,"Estimate"]
14 - 1.96*fitsummary$coefficients[2,"Std.err"], methods = "CC_adjusted")
15
16 # marginal effects: margi_df <- emmeans(fit, "ttt") %>% as.data.table()
17
18 margi_df[,methods := "CC_adjusted"] setnames(margi_df,"emmean","LDA")
19 margi_df[,LDA_inf := LDA - 1.96*SE] margi_df[,LDA_sup := LDA + 1.96*SE]
20 margi_df[,LDA_var := SE^2]
21
22 output <- rbind(diff,margi_df[,.(ttt,LDA,LDA_sup,LDA_inf,LDA_var,methods)])
23
24 return(list(output = output,fit = fit)) }
25
26
27
28
29
30 # Not adjusted complete case imputation
31
32
33
34
35 ```{r setup, include=FALSE}
36
37 export("CC_raw")
38 CC_raw <- function(datain,
39 # data
40 var = "CDAI_fu",
41 # variable measuring effectiveness
42 thres = 10,
43 # threshold for remission or LDA
44 ttt_var = "ttt",
45 ref_ttt = "ttt_ref")
46 # variable name containing the treatment
47 {
48 data <- check_data(datain,var,ttt,ref_ttt)
49 # raw proportion
50 raw_prop <- data[!is.na(get(var)),
51 .(LDA = sum(get(var)<=thres)/.N,
52 methods = "CC_raw",
53 N = .N),
54 by = ttt]
55
56
57
58
59
60

```

```

1
2
3   # calculation of the Standard error
4   raw_prop[,c("LDA_inf","LDA_sup") := lapply(c(-1.96,1.96),function(z)
5   {
6     LDA + z*sqrt(LDA*(1-LDA)/N)
7     }])
8
9   # difference between treatments
10  diff_tmp <- raw_prop[,.(ttt = "diff",
11                          LDA = LDA[ttt == "ttt_1"]-LDA[ttt ==
12 "ttt_ref" ],
13                          methods = methods[1] ,
14                          SE = (sum(1/N))/2 + 1.96*sqrt(sum( LDA*(1-LDA)/N
15 )))])
16
17
18  diff_tmp[,LDA_inf := LDA - SE]
19  diff_tmp[,LDA_sup := LDA + SE]
20
21  # bind outputs
22  output <- rbind(diff_tmp[,.(ttt,LDA,LDA_inf,LDA_sup,methods)],
23                 raw_prop[,-"N"])
24
25  # change name back
26  setnames(output,"ttt",ttt_var)
27  return(output)
28 }
29
30

```

Adjusted complete case imputation

```

31
32
33 ````{r setup, include=FALSE} export("CC_adjusted") CC_adjusted = function(datain, var =
34 "CDAI_fu", thres = 10, ttt_var = "ttt", ref_ttt = "ttt_ref", covariates =
35 c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAIO") ) # variable
36 name containing the treatment { data <- check_data(datain,var,ttt_var,ref_ttt) data[,LDA :=
37 get(var) <= thres] output <- adjusted_model(data = data[!is.na(get(var))], covariates =
38 covariates)$output
39
40

```

```

41 output[,methods := "CC_adjusted"] # change name back setnames(output,"ttt",ttt_var)
42 return(output) }
43
44

```

```

45
46 # LOCF imputation
47
48

```

```

49
50 ````{r setup, include=FALSE}
51 export("LOCF")
52

```

```

53 LOCF <- function(datain,
54                  var = "CDAI_fu",
55                  var_before = "CDAI_beforefu",
56
57
58
59

```

```

1
2
3         thres = 10,
4         ttt_var = "ttt",
5         ref_ttt = "ttt_ref",
6         covariates =
7 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0")
8 ) {
9   data <- copy(datain)
10
11   data <- check_data(datain, var, ttt_var, ref_ttt)
12   data[is.na(get(var)), c(var) := get(var_before)]
13   data[,LDA := get(var) <= thres]
14
15
16   output <- adjusted_model(data = data,
17                             covariates = covariates)$output
18
19   output[,methods := "LOCF"]
20   # change name back
21   setnames(output, "ttt", ttt_var)
22   return(output)
23 }
24
25
26

```

Lundex imputation

```

27
28
29 ``{r setup, include=FALSE} export("Lundex") Lundex <- function(datain, var = "CDAI_fu",
30 thres = 10, ttt_var = "ttt", ref_ttt = "ttt_ref", treatment_duration = "treatment_duration",
31 stop_var = "stopany", covariates =
32 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0"), boot_num =
33 1000) {
34
35   data <- check_data(datain, var, ttt_var, ref_ttt) data[,LDA := get(var) <= thres] ####
36   bootstrap for SE data[,tmp := 1] # replicated data for bootstrap replicateddata <-
37   data[CJ(tmp = 1, boot = 1:boot_num), on = "tmp", allow.cartesian=TRUE] # sample with
38   replacement for each boot sampled_idx <- replicateddata[,I[sample(1:N, replace = T)], by =
39   boot]$V1 bootstrapdata <- replicateddata[sampled_idx]
40
41
42   # raw proportions raw_prop <- bootstrapdata[!is.na(get(var)), { adjusted_model(data
43   = .SD)$output %>% .[ttt != "diff", (ttt, LDA_raw = LDA) ] }, by = .(boot)]
44
45   # surv analysis for each bootstrapped dataset surv_formula <-
46   as.formula(paste0("Surv(", treatment_duration, ", ", stop_var, ") ~ ttt"))
47
48   surv_coeff <- bootstrapdata[, { temp.km <- survfit(surv_formula, data = .SD) list(surv =
49   summary(temp.km, times = 1)$surv, ttt = gsub("ttt=", "", unique(summary(temp.km)
50   $strata))) }, by = boot]
51
52
53   # LDA: LDA raw * surv coeff tmp_bootstrap <- merge(raw_prop, surv_coeff, by =
54   c("boot", "ttt")) tmp_bootstrap[,LDA := LDA_raw*surv]
55
56
57
58
59

```

```

1
2
3 # difference between treatments diff_boot <- tmp_bootstrap[,.(ttt = "diff", LDA = LDA[ttt !=
4 ref_ttt] - LDA[ttt == ref_ttt]), by = boot]
5
6 tot_bottstrap <- rbind(diff_boot[,.(ttt, LDA, boot)], tmp_bootstrap[,.(ttt, LDA, boot)])
7
8 # calculate the mean and the SE: output <- tot_bottstrap[,.(LDA = mean(LDA), LDA_sup =
9 quantile(LDA, 0.975), LDA_inf = quantile(LDA, 0.025) ), by = ttt] # change name back
10 output[,methods := "LUNDEX"]
11
12 setnames(output,"ttt",ttt_var) return(output) }
13
14
15
16 # non-responder imputation
17
18
19
20 ```{r setup, include=FALSE}
21 export("NRI")
22 NRI = function(datain,
23               var="CDAI_fu",
24               thres = 10,
25               ttt_var = "ttt",
26               ref_ttt = "ttt_ref",
27               covariates =
28 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0")
29 )
30 # variable name containing the treatment
31 {
32   data <- check_data(datain,var,ttt_var,ref_ttt)
33   data[,LDA := get(var) <= thres]
34   data[is.na(LDA),LDA := 0] # missing are non responders
35   output <- adjusted_model(data = data,
36                             covariates = covariates)$output
37
38   # change name back
39   setnames(output,"ttt",ttt_var)
40   output[,methods := "NRI"]
41   return(output)
42 }
43
44
45
46

```

Inverse probability weighting imputation

```

47
48
49 ```{r setup, include=FALSE} export("IPW") IPW <- function(datain, var = "CDAI_fu", thres =
50 10, ttt_var = "ttt", ref_ttt = "ttt_ref", counfounders =
51 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0"), attrition =
52 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0", "stopreason")) {
53
54   data <- check_data(datain, var, ttt_var, ref_ttt, othervar = c(counfounders,attrition))
55
56
57
58
59
60

```

```

1
2
3 data[,ttt2 := as.numeric(ttt != ref_ttt)] # weight for confounding formula_coeff <-
4 paste0("~",paste0(counfunders,collapse = "+")) function_call <- paste0('IPWT <-
5 ipwpoint( exposure = ttt2 , family = "binomial", link = "logit", numerator = ~ 1,
6 denominator =',formula_coeff,', data = data, trunc = 0.01 )') eval(parse(text = function_call))
7 datasw<- IPW Tipw.weights
8
9
10 # weights for attrition formula_attr <- paste0("~",paste0(attrition,collapse = "+"))
11 data[,MISS := as.numeric(is.na(get(var)))] function_call <- paste0('IPCT <-
12 ipwpoint( exposure = MISS , family = "binomial", link = "logit", numerator = ~ 1,
13 denominator =',formula_attr,', data = data )') eval(parse(text = function_call)) data
14 swc<- IPC Tipw.weights
15
16 dataNoNA <- na.omit(data[,.(ttt,get(var),sw,swc,ID_ttt) %>%
17 setNames(c("ttt",var,"sw","swc","ID_ttt"))]) dataNoNA[,LDA := as.numeric(get(var) <=
18 thres)]
19
20
21 output <- adjusted_model(data = dataNoNA, weights = dataNoNA$sw*dataNoNA$swc)
22 $output
23
24 output[,methods := "IPW"]
25
26 # change name back setnames(output,"ttt",ttt_var) return(output)
27
28 }
29
30
31
32 # Confounder-Adjusted Response Rate with Attrition Correction (CARRAC)
33 imputation
34
35
36 ```{r setup, include=FALSE}
37 export("CARRAC")
38 CARRAC <- function(datain,
39                     var = "CDAI_fu",
40                     thres = 10,
41                     ttt_var = "ttt",
42                     ref_ttt = "ttt_ref",
43                     counfunders =
44                     c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAI0"),
45                     attrition =
46                     c("Disease_duration","concomitantCsDMARD",
47                       "Prev_bDMARD3","CDAI0","stopreason"),
48                     seed = NA) {
49
50   data <- check_data(datain,var,ttt_var,ref_ttt)
51   dataS <- data[,.SD,.SDcols =
52   c("ID_ttt",var,"ttt",union(counfunders,attrition))]
53
54
55
56
57
58
59
60

```

```

1
2
3 impute_data <- mice(
4   dataS,
5   m = 10,
6   method = "pmm",
7   maxit = 5,
8   printFlag = F, seed = seed
9 )
10
11 # open the data
12 impute_data_complete <- setDT(complete(impute_data, action = "long"))
13 # calculate LDA
14 impute_data_complete[,LDA := get(var) <= thres]
15
16 # get LDA and error for each imputation
17 res_mice <- lapply(seq(1:impute_data$m), function(imp){
18
19   adjusted_model(data = impute_data_complete[.imp == imp],
20                 covariates = counfunders)$output
21
22 }) %>% rbindlist()
23
24 res_mice_2 <- lapply(seq(1:impute_data$m), function(imp){
25
26   adjusted_model(data = impute_data_complete[.imp == imp],
27                 covariates = counfunders)$fit
28
29 })
30
31 test <- pool(res_mice_2)
32 df_pval <- summary(test) %>% as.data.table()
33 p.output <- df_pval[grepl("ttt", term), p.value]
34
35
36 # pooling
37 pool_res <- res_mice[,.(
38   LDA_mi = mean(LDA),
39   w = mean(LDA_var),
40   m = .N,
41   b = 1/(.N-1)*sum( (LDA-mean(LDA))^2 )
42   ),by = ttt]
43
44 pool_res[,LDA_var := w + (1+1/m)*b]
45 pool_res[,LDA_sd := sqrt(LDA_var)]
46
47 # mean, 95% CI
48 output <- pool_res[,.(ttt,
49                       LDA_mi,
50                       LDA_mi + 1.96*LDA_sd,
51                       LDA_mi-1.96*LDA_sd) %>%
52                       setNames(c("ttt", "LDA", "LDA_sup", "LDA_inf"))]
53
54
55 output[,methods := "CARRAC"]
56
57
58
59
60

```

```
1
2
3     output[ttt == "diff",p := p.output]
4
5     # change name back
6     setnames(output,"ttt",ttt_var)
7
8     return(output)
9 }
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

For peer review only

3 - FINAL FIGURES CODE

10/11/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)
library(patchwork) # package to compose multiplots !
library(ggpubr)
library(grid)

rm(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))

load("./3_clean_output/full_workspaces/workspace_1.RData")
load("./3_clean_output/full_workspaces/workspace_2.RData")
load("./3_clean_output/full_workspaces/workspace_3.RData")
```

1. Common theme

```
theme_benoit = function(){
  theme_pubclean()+
    theme(panel.grid.major.x = element_line(linetype = "dotted", colour
= "grey50"),
          panel.grid.major.y = element_line(linetype = "dotted", colour
= "grey50"),
          axis.title.y = element_text(margin = margin(r = .2, unit =
"cm")),
          axis.title.x = element_text(margin = margin(t = .2, unit =
```

```
1  
2  
3 "cm")),  
4 plot.title = element_text(margin = margin(b = .5, unit =  
5 "cm")))  
6 }  
7  
8  
9
```

1. [0] Mini explanation

TC lenght

```
13 BARI_DATA[,time_on_drug_year := time_on_drug/365.25]
```

```
15  
16 p1 <- ggplot(BARI_DATA)+  
17   geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =  
18   1/12)+
```

```
19  
20   scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+  
21   labs(x = "Duration of observation (years)",  
22        y = "Number of TC",  
23        title = "Time of observation for all included TC")+  
24   ylim(-25,NA)+  
25   theme_benoit()
```

```
26 p1
```

```
27  
28  
29 p2 <- ggplot(BARI_DATA)+  
30   geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =  
31   "grey80")+  
32   theme_void()
```

```
33  
34 plot_mini_exploration <- p1 + inset_element(p2,0.01,0.05,0.99,0.2)  
35 plot_mini_exploration
```

Saving plot

```
36  
37  
38  
39 png("./3_clean_output/figures/PL0T_Exploration_TC_duration.png",  
40     width = 7,  
41     height = 5,  
42     units = "in",  
43     res = 300) # opening graphic device  
44 plot_mini_exploration  
45 dev.off() # closing graphic device
```

TC lenght for BARI only

```
46  
47  
48  
49 data_sub <- BARI_DATA[cohort == "BARI"]
```

```
50  
51  
52 p1 <- ggplot(data_sub)+  
53   geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =  
54   1/13, fill = "red3")+
```

```
55  
56   scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
```

```
1
2
3     labs(x = "Duration of observation (years)",
4           y = "Number of TC",
5           title = "A - BARI")+
6     ylim(-11,50)+
7     theme_benoit()
8 p1
9
10
11 p2 <- ggplot(data_sub)+
12     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
13 "grey80")+
14     theme_void()
15
16 plot_mini_exploration_bari <- p1 +
17 inset_element(p2,0.01,0.05,0.99,0.2)
18 plot_mini_exploration_bari
19
20 TC lenght for TNFi only
21
22 data_sub <- BARI_DATA[cohort == "TNFi"]
23
24 p1 <- ggplot(data_sub)+
25     geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
26 1/13, fill = "green2")+
27
28     scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
29     labs(x = "Duration of observation (years)",
30          y = "Number of TC",
31          title = "B - TNFi")+
32     ylim(-11,50)+
33     theme_benoit()
34 p1
35
36
37 p2 <- ggplot(data_sub)+
38     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
39 "grey80")+
40     theme_void()
41
42
43 plot_mini_exploration_tnfi <- p1 +
44 inset_element(p2,0.01,0.05,0.99,0.2)
45 plot_mini_exploration_tnfi
46
47 TC lenght for OMA only
48
49 data_sub <- BARI_DATA[cohort == "OMA"]
50
51 p1 <- ggplot(data_sub)+
52     geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
53 1/13, fill = "blue2")+
54
55     scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
56
57
58
59
60
```

```

1
2
3     labs(x = "Duration of observation (years)",
4           y = "Number of TC",
5           title = "C - OMA")+
6     ylim(-11,50)+
7     theme_benoit()
8 p1
9
10
11 p2 <- ggplot(data_sub)+
12     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
13 "grey80")+
14     theme_void()
15
16 plot_mini_exploration_oma <- p1 + inset_element(p2,0.01,0.05,0.99,0.2)
17 plot_mini_exploration_oma
18
19 multiplot
20
21 multi_plot <- plot_mini_exploration_bari + plot_mini_exploration_tnfi
22 + plot_mini_exploration_oma
23 multi_plot
24
25 median(BARI_DATA[cohort == "BARI", time_on_drug])
26 median(BARI_DATA[cohort == "TNFi", time_on_drug])
27 median(BARI_DATA[cohort == "OMA", time_on_drug])
28
29 Saving plot
30
31 png("./3_clean_output/figures/
32 PLOT_Exploration_TC_duration_3_groups.png",
33     width = 9,
34     height = 5,
35     units = "in",
36     res = 300) # opening graphic device
37 multi_plot
38 dev.off() # closing graphic device
39
40
41
42
43
44

```

1. [1] Survival analysis

Forest plot BARI vs TNFi + BARI vs OMA

```

45 meanall <- summary(BARI1.adj.mi)$coefficients[1:14,"exp(coef)"]
46 lowerall <- summary(BARI1.adj.mi)$conf.int[1:14,"lower .95"]
47 upperall <- summary(BARI1.adj.mi)$conf.int[1:14,"upper .95"]
48 textall <- c("Treatment (vs BARI)", "Age (decades)", "BMI",
49 "Concomitant csDMARD", "Concomitant glucocorticoid", "CDAI score (10
50 pts)", "Disease duration (decades)", "Current smoker (vs non-smoker)",
51 "Former smoker (vs non-smoker)", "2nd line therapy (vs 1st)", "3rd
52 line therapy (vs 1st)", "4th or later line (vs 1st)", "Female gender",
53 "Seropositivity (RF or ACPA)")
54
55
56
57
58
59

```

```

1
2
3 dfall1 <- data.table(textall, meanall, lowerall, upperall)
4 dfall1[,ttt := "TNFi"]
5
6 meanall <- summary(BARI2.adj.mi)$coefficients[1:14,"exp(coef)"]
7 lowerall <- summary(BARI2.adj.mi)$conf.int[1:14,"lower .95"]
8 upperall <- summary(BARI2.adj.mi)$conf.int[1:14,"upper .95"]
9 dfall2 <- data.table(textall, meanall, lowerall, upperall)
10 dfall2[, ttt := "OMA"]
11
12
13 dfall <- rbind(dfall1,dfall2)
14 dfall$textall <- factor(dfall$textall, levels = rev(textall))
15
16 text_high <- textGrob("\u2192 Reduces \ndrug maintenance",
17 gp=gpar(fontsize=8, fontface="bold"))
18 text_low <- textGrob("\u2190 Improves \ndrug maintenance",
19 gp=gpar(fontsize=8, fontface="bold"))
20
21
22 HR_plot <- ggplot(data=dfall,
23 aes(x = textall,
24 y = meanall,
25 ymin = lowerall,
26 ymax = upperall,
27 color = ttt))+
28 geom_hline(yintercept =1, linetype=2)+
29 geom_point(size=2,position = position_dodge(width = .7))+
30 geom_errorbar(position = position_dodge(width = .7))+
31 labs(x = "",y = "",color = "")+
32 scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
33
34 +
35 theme(axis.line.x = element_line(size = 0.5),
36 axis.text = element_text(face = "bold", color = "black"),
37 legend.position="top",
38 legend.key = element_blank(),
39 plot.margin = unit(c(1,3,2,1),"lines"))+
40 coord_flip(clip = "off")+
41 annotation_custom(text_high,
42 xmin=-0.64,xmax=-0.64,ymin=.2,ymax=.2)+
43 annotation_custom(text_low,
44 xmin=-0.64,xmax=-0.64,ymin=-.15,ymax=-.15)+
45 theme_pubclean()+
46 scale_color_manual(breaks = c("OMA","TNFi"),values =
47 c("blue3","green3"),labels = c("OMA","TNFi"))
48

```

HR_plot

Saving the plot in PNG file

```
png("./3_clean_output/figures/PLOT FOREST BARI vs TNFi vs OMA HR.png",
```

```
width = 7,
```

```

1
2
3     height = 5.5,
4     units = "in",
5     res = 300)
6
7 HR_plot
8
9
10 dev.off() # closing graphic device
11
12 BARI vs TNFi
13
14 Non-adjusted Kaplan-Meier curves
15
16 BARI vs TNFi
17
18 BARI1[,time_on_drug_year := time_on_drug/365.25]
19
20 surv_object1 <- Surv(time = BARI1$time_on_drug_year, event =
21 BARI1$stop_DMARD) # indicate time on drug and stop variable
22 fit1 <- survfit(surv_object1 ~ cohort, data = BARI1)
23
24 survplot_1 <- ggsurvplot(fit1, data = BARI1, # plot
25                          pval = T,
26                          pval.method = TRUE,
27                          legend.title = "Groups :",
28                          legend.labs = c("BARI", "TNFi"),
29                          xlab = "Time (years)",
30                          xlim = c(0, 2.5),
31                          censor = FALSE,
32                          title = "Non-adjusted model of drug
33 discontinuation \nby type of treatment",
34                          surv.median.line = "v",
35                          linetype = 1,
36                          size = 1.5,
37                          #palette = c("grey78", "grey10"),
38                          palette = c("red3", "green2"), # pour mettre
39 les couleurs
40
41                          ggtheme = theme_benoit(),
42                          risk.table = T)
43
44
45 values <- summary(fit1)$table[,"median"]
46 df <- data.frame(y = .1,x = values+.2,label =
47 as.character(round(values,2)))
48
49 survplot_1$plot <- survplot_1$plot +
50   geom_text(data = df,aes(x,y,label = label), color = c("red3",
51 "green2"), size = 5)
52
53
54
55 print(survplot_1)
56
57
58
59
60

```

1
2
3 Saving surplot
4

5 png("./3_clean_output/figures/PLOT BARI vs TNFi curves non adjusted
6 COLOR.png",
7 width = 7,
8 height = 7, units = "in",
9 res = 300) # opening graphic device
10 survplot_1
11

12 dev.off() # closing graphic device
13

14 Saving the plot curv object for Lilly
15

16 plot_BARI_vs_TNFi_data <- survplot_1\$data.survplot
17 write.xlsx(plot_BARI_vs_TNFi_data, file =
18 "./3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_non_adjust
19 ed.xlsx", row.names = F)
20

21 **Home-made attempt to obtain adjusted curves based on imputed data**

22 dummy_cox_impute1 <- mice::complete(imputed_data1, "long", include =
23 T)
24 dummy_cox_impute1 <- dummy_cox_impute1[dummy_cox_impute1\$.imp != 0,]
25 dummy_cox_impute1\$time_on_drug_year <-
26 dummy_cox_impute1\$time_on_drug/365.25
27
28

29 BARI_fit1 <- survfit(coxph(Surv(time = time_on_drug_year, event =
30 stop_DMARD) ~ cohort+
31 I(age_base/10)+
32 bmi_base+
33 TC_with_csDMARD+
34 PREDNISON_STEROID+
35 CDAI0+
36 I(disease_duration_base_years/10)+
37 C(smoker_base, base=3)+
38 line_of_therapy+
39 gender+
40 seropositivity_base+
41 cluster(patient_id)+
42 strata(cohort), dummy_cox_impute1), data =
43
44 dummy_cox_impute1)
45
46

47 survplot_1_adj <- ggsvplot(BARI_fit1, data = dummy_cox_impute1,
48 variable = "cohort",
49
50 xlab = "Time (years)",
51 title = "A - BARI vs TNFi",
52 legend.title = "Groups :",
53 legend.labs = c("BARI", "TNFi"),
54 censor = FALSE,
55 xlim = c(0, 2.5),
56
57
58
59
60

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

```

```

surv.median.line = "v",
linetype = 1,
size = 1.5,
ggtheme = theme_benoit(),
# palette = c("grey78", "grey10")
palette = c("red2", "green3" )+
  labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,\n
concomitant glucocorticoid, baseline CDAI, disease duration
(decades),\n smoking status, line of therapy, gender, seropositivity")

# adding days label
values <- summary(BARI_fit1)$table[,"median"]
df <- data.frame(y = .1,x = values+.1,label =
as.character(paste(round(values*365.25,2), "\n days")))
df[1,2] <- 1.82

survplot_1_adj$plot$labels$y <- "Proportion still on drug" # to change
the label

survplot_1_adj$plot <- survplot_1_adj$plot +
  geom_text(data = df,aes(x,y,label = label), color = c("red3",
"green3"), size = 5)

# adding HR et p val label
HR <- data.frame(y = 0.1, x = 0.5, label = paste("HR =",
round(exp(BARI1.adj.mi$coefficients[1]), 2), "\n", "p =",
round(summary(BARI1.adj.mi)$coefficients[1,"Pr(>|z|)"], 4) ) )

survplot_1_adj$plot <- survplot_1_adj$plot +
  geom_text(data = HR,aes(x,y,label = label) , size = 5)

# final print
survplot_1_adj

Saving the survival plot in PNG file

png("./3_clean_output/figures/PLOT BARI vs TNFi curves adjusted.png",

width = 7,
height = 5,
units = "in",
res = 300)

survplot_1_adj

dev.off() # closing graphic device

```

Sensitivity analysis (RiskRegression Package)

First plot to get the difference in average treatment effect in percentage


```
1
2
3 dt.out$time_years <- dt.out$time/365.25
4
5 plot.ate.diff <- ggplot(dt.out[type == "meanRisk"], aes(x =
6 time_years, group = level))+
7   geom_vline(xintercept = 1, linetype = 2, size = 1, color = "grey20")
8 +
9   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
10 0.4)+
11   geom_line(aes(y = estimate, color = level), size = 1)+
12
13   scale_colour_manual(values = c("red2","green3"))+
14   scale_fill_manual(values = c("red2","green3"))+
15   theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
16 legend.position="top" )+
17   scale_x_continuous(breaks=seq(0,2.5,0.25)) +
18   scale_y_continuous(labels = scales::percent, limits = c(0,0.65))+
19
20   xlab("Years since intiation of treatment")+
21   ylab("Absolute Risk of treatment discontinuation (%)")+
22   labs(colour="Groups:", fill = "Groups:", title = "A - BARI vs TNFi")
23 +
24   labs(group = "Groups:")+
25   theme_benoit()+
26   theme(axis.title.x = element_text(margin = margin(t = .3,unit =
27 "cm")),
28 axis.title.y = element_text(margin = margin(r = .3,unit =
29 "cm")))
30
31 plot.ate.diff
32
33 Saving Plot
34
35 png("./3_clean_output/figures/PLOT BARI vs TNFi curves AIPTW.png",
36 width = 1300, height = 650, res = 120) # opening graphic device
37
38 plot.ate.diff
39
40 dev.off() # closing graphic device
41
42 Second plot to get the ratio in average treatment effect
43
44 plot.ate.ratio <- ggplot(dt.out[type == "ratioRisk"], aes(x = time,
45 group = level))+
46   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
47 0.4)+
48   geom_line(aes(y = estimate, color = level), size = 2)+
49
50   theme_benoit()+
51   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
52   scale_x_continuous(breaks=seq(0,500,50))+
53
54
55
56
57
58
59
60
```

```
1
2
3     scale_y_continuous(limits = c(0.9,4.5))+
4
5     xlab("Days since intiation of treatment")+
6     ylab("Ratio in Average Treatment Effect")+
7     labs(colour="treatment", fill = "treatment")
8
9
10 plot.ate.ratio
11
12 BARI vs OMA
13
14 Non-adjusted Kaplan-Meier curves
15
16 BARI vs OMA
17
18 BARI2[,time_on_drug_year := time_on_drug/365.25]
19
20 surv_object2 <- Surv(time = BARI2$time_on_drug_year, event =
21 BARI2$stop_DMARD)
22 fit2 <- survfit(surv_object2 ~ cohort, data = BARI2) # this function
23 creates the data for Kaplan Meyer
24 survplot_2 <- ggsurvplot(fit2, data = BARI2, # plot
25     pval = T,
26     pval.method = TRUE,
27     legend.title = "Groups :",
28     legend.labs = c("BARI", "OMA"),
29     xlab = "Time (days)",
30     xlim = c(0, 2.5),
31     censor = FALSE,
32     title = "Non-adjusted model of drug discontinuation by type
33 of treatment",
34     surv.median.line = "v",
35     linetype = 1,
36     size = 1.5,
37     ggtheme = theme_benoit(),
38     #palette = c("grey78", "grey50"),
39     palette = c("red3", "blue2"), # to put colors
40     risk.table = T)
41
42 survplot_2
43
44 Saving surplot
45
46 png("./3_clean_output/PLOT BARI vs OMA curves non adjusted COLOR.png",
47 width = 1000, height = 600, res = 100) # opening graphic device
48
49 survplot_2
50
51 dev.off() # closing graphic device
52
53 Saving the plot curv object for Lilly
54
55
56
57
58
59
60
```

```

1
2
3 plot_BARI_vs_OMA_data <- survplot_2$data.survplot
4 write.xlsx(plot_BARI_vs_OMA_data, file =
5 ".\3_clean_output\Lilly_curves_excel\plot_BARI_vs_OMA_data_non_adjuste
6 d.xlsx", row.names = F)
7

```

Home-made attempt to obtain adjusted cuves based on imputed data :

```

8
9 dummy_cox_impute2 <- mice::complete(imputed_data2, "long", include =
10 T)
11 dummy_cox_impute2 <- dummy_cox_impute2[dummy_cox_impute2$.imp != 0,]
12 dummy_cox_impute2$time_on_drug_year <-
13 dummy_cox_impute2$time_on_drug/365.25
14
15
16 BARI_fit2 <- survfit(coxph(Surv(time = time_on_drug_year, event =
17 stop_DMARD) ~ cohort+
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
          I(age_base/10)+
          bmi_base+
          TC_with_csDMARD+
          PREDNISON_STEROID+
          CDAI0+
          I(disease_duration_base_years/10)+
          C(smoker_base, base=3)+
          line_of_therapy+
          gender+
          seropositivity_base+
          cluster(patient_id)+
          strata(cohort),dummy_cox_impute2), data =
dummy_cox_impute2)

survplot_2_adj <- ggsurvplot(BARI_fit2, data = dummy_cox_impute2,
variable = "cohort",
  xlab = "Time (years)",
  title = "B - BARI vs OMA",
  legend.title = "Groups :",
  legend.labs = c("BARI", "OMA"),
  censor = FALSE,
  xlim = c(0, 2.5),
  surv.median.line = "v",
  linetype = 1,
  size = 1.5,
  ggtheme = theme_benoit(),
  # palette = c("grey78", "grey50")
  palette = c("red2", "blue3") # to change colors
)+
  labs(caption = "Adjusted for : age, BMI, concomitant csDMARD, \n
concomitant glucocorticoid, baseline CDAI, disease duration
(decades),\n smoking status, line of therapy, gender, seropositivity")

# adding Days label
values <- summary(BARI_fit2)$table[,"median"]
df <- data.frame(y = .1,x = values+.1,label =

```

```
1
2
3 as.character(paste(round(values*365.25,2), "\n days"))
4 df[1,2] <- 1.82
5
6 survplot_2_adj$plot$labels$y <- "Proportion still on drug" # to change
7 the label
8
9
10 survplot_2_adj$plot <- survplot_2_adj$plot +
11   geom_text(data = df,aes(x,y,label = label), color = c("red3",
12 "blue2"), size = 5)
13
14 # adding HR et pval label
15 HR <- data.frame(y = 0.1, x = 0.5, label = paste("HR =",
16 round(exp(BARI2.adj.mi$coefficients[1]), 2), "\n", "p =",
17 round(summary(BARI2.adj.mi)$coefficients[1,"Pr(>|z|)"], 4) ) )
18
19 survplot_2_adj$plot <- survplot_2_adj$plot +
20   geom_text(data = HR,aes(x,y,label = label) , size = 5)
21
22
23
24
25 # final print
26 survplot_2_adj
27
28 summary(BARI_fit2, times = 1) # to see detailed surv probabilities at
29 given timepoints
30
31 Saving the survival plot in PNG file
32
33 png("./3_clean_output/PLOT BARI vs OMA curves adjusted.png", width =
34 1000, height = 600, res = 100) # opening graphic device
35
36 survplot_2_adj
37
38 dev.off() # closing graphic device
39
40
41 Sensitivity analysis (RiskRegression package)
42
43 First plot to get the difference in average treatment effect in percentage
44
45 dt.out2$time_years <- dt.out2$time/365.25
46
47 plot.ate.diff2 <- ggplot(dt.out2[type == "meanRisk"], aes(x =
48 time_years, group = level))+
49   geom_vline(xintercept = 1, linetype = 2, size = 1, color = "grey20")
50 +
51   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
52 0.3)+
53   geom_line(aes(y = estimate, color = level), size = 1)+
54
55 theme_benoit() + theme(legend.spacing.x = unit(0.2, 'cm'),
56
57
58
59
```

```

1
2
3 legend.position="top" )+
4   scale_x_continuous(breaks=seq(0,2.5,0.25)) +
5   scale_y_continuous(labels = scales::percent, limits = c(0,0.65))+
6
7   xlab("Years since initiation of treatment")+
8   ylab("Absolute Risk of treatment discontinuation (%)")+
9   labs(colour="Groups:", fill = "Groups:", title = "B - BARI vs OMA")+
10  labs(group = "Groups:")+
11
12
13   scale_colour_manual(values = c("red2","blue3"))+
14   scale_fill_manual(values = c("red2","blue3"))
15

```

```
16 plot.ate.diff2
```

```
17 Saving Plot
```

```
18
19 png("./3_clean_output/PLOT BARI vs OMA curves AIPTW.png", width =
20 1300, height = 650, res = 120) # opening graphic device
21
22

```

```
23 plot.ate.diff2
```

```
24 dev.off() # closing graphic device
25

```

```
26 Second plot to get the ratio in average treatment effect
27

```

```
28
29 plot.ate.ratio2 <- ggplot(dt.out2[type == "ratioRisk"], aes(x = time,
30 group = level))+
31   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
32 0.3)+
33   geom_line(aes(y = estimate, color = level), size = 1)+
34
35   theme_benoit()+
36   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
37   scale_x_continuous(breaks=seq(100,400,50))+
38   scale_y_continuous(limits = c(0.8,3))+
39
40
41   xlab("Days since initiation of treatment")+
42   ylab("Ratio in Average Treatment Effect")+
43   labs(colour="treatment", fill = "treatment")
44

```

```
45 plot.ate.ratio2
46

```

47 Multipanel plots

```
48 To update using patchwork
```

```
49 For the paper Non adjuted curves
```

```
50 # Creating list object
```

```
51 plots <- list()
52
53
54
55
56
57
58
59
60

```

```
1
2
3   plots[[1]] <- survplot_1
4   plots[[2]] <- survplot_2
5
6   # Nice function
7
8   multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
9
10  Option 2 putting all data on one panel Kaplan Meier
11
12  BARI vs TNFi vs OMA
13
14  BARI_DATA[,time_on_drug_year := time_on_drug/365.25]
15
16  surv_object3 <- Surv(time = BARI_DATA$time_on_drug_year, event =
17  BARI_DATA$stop_DMARD)
18  fit3 <- survfit(surv_object3 ~ cohort, data = BARI_DATA) # this
19  function creates the data for Kaplan Meyer
20  survplot_3 <- ggsurvplot(fit3, data = BARI_DATA, # plot
21  pval = F,
22  pval.method = F,
23  legend.title = "Groups",
24  legend.labs = c("BARI", "TNFi", "OMA"),
25  xlab = "Time (years)",
26  xlim = c(0, 2.5),
27  censor = FALSE,
28  # title = "Non-adjusted drug discontinuation by type of
29  treatment (Kaplan-Meier)",
30  surv.median.line = "v",
31  linetype = 1,
32  size = 1.5,
33  ggtheme = theme_benoit(),
34  palette = c("red3", "green2", "blue2"), # to put colors
35  risk.table = T)
36
37
38
39  values <- summary(fit3)$table[, "median"]
40  df <- data.frame(y = .1, x = values+.1, label =
41  as.character(paste(round(values*365.25,2), "\n days")))
42  df[3,2] <- 1.72
43
44  survplot_3$plot <- survplot_3$plot +
45  geom_text(data = df, aes(x,y,label = label), color = c("red3",
46  "green2", "blue2"), size = 5)
47
48
49  survplot_3$plot$labels$y <- "Proportion still on drug" # to change the
50  label
51  survplot_3
52
53  Saving survplot
54
55  png("./3_clean_output/figures/PLOT BARI vs TNFi vs OMA curves non
56  adjusted COLOR.png", width = 800, height = 600, res = 100) # opening
57
58
59
60
```

```
1
2
3   graphic device
4
5   survplot_3
6
7   dev.off() # closing graphic device
8
9   Adjusted curves
10
11  # Creating list object
12
13  plots <- list()
14  plots[[1]] <- survplot_1_adj
15  plots[[2]] <- survplot_2_adj
16
17  # Nice function
18
19
20  multi_plot_cox <- arrange_ggsurvplots(plots, print = T, ncol = 2)
21
22  png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA curves
23  adjusted COLOR.png", width = 1000, height = 600, res = 100) # opening
24  graphic device
25
26  multi_plot_cox
27
28  dev.off() # closing graphic device
29
30  All curves
31
32  # Creating list object
33
34  plots <- list()
35  plots[[1]] <- survplot_1
36  plots[[3]] <- survplot_2
37  plots[[2]] <- survplot_1_adj
38  plots[[4]] <- survplot_2_adj
39
40
41  # Nice function
42
43
44  multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2, nrow =
45  2)
46
47  # but does not display properly now.. :(
48
49  AIPTW absolute risk of treatment discontinuation biplot
50
51  plot.ate.diff + plot.ate.diff2
52
53  png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA AIPTW curves
54  adjusted COLOR.png", width = 1000, height = 600, res = 100) # opening
55  graphic device
56
57
58
59
60
```

```
1
2
3 plot.ate.diff + plot.ate.diff2
4
5 dev.off() # closing graphic device
6
```

7 Diagnostic multipanel plots

8 Asked by Lilly statistician to show balance in this analysis.

```
9
10
11 pscore_plot <- ggplot(test.data, aes(x = pscores, color = cohort, fill
12 = cohort)) +
13   geom_density(alpha = .47) +
14
15   theme_minimal()+
16   theme(axis.ticks.y = element_blank(),
17         panel.grid.minor = element_blank(),
18         legend.title = element_blank(),
19         text = element_text(size = 16),
20         axis.title.x = element_text(hjust = 0.2, size = 16))+
21
22   scale_colour_manual(values = c("red2","green3"))+
23   scale_fill_manual(values = c("red2","green3"))+
24
25   xlab("Probability of being assigned BARI or TNFi") +
26   ylab("Density") +
27   labs(title = "A1 - BARI vs TNFi")
28
29
```

```
30 pscore_plot # overlap
```

```
31
32 pscore_plot2 <- ggplot(test.data2, aes(x = pscores, color = cohort,
33 fill = cohort)) +
34   geom_density(alpha = .47) +
35
36   theme_minimal()+
37   theme(axis.ticks.y = element_blank(),
38         panel.grid.minor = element_blank(),
39         legend.title = element_blank(),
40         text = element_text(size = 16),
41         axis.title.x = element_text(hjust = 0.2, size = 16))+
42
43   scale_colour_manual(values = c("red2","blue3"))+
44   scale_fill_manual(values = c("red2","blue3"))+
45
46   xlab("Probability of being assigned BARI or OMA") +
47   ylab("Density") +
48   labs(title = "A2 - BARI vs OMA")
49
```

```
50
51
52 pscore_plot2
53 # Good overlap
54
55
56
57
58
59
```



```
1
2
3 library(cobalt)
4
5 # BARI vs TNFi
6 B1 <- love.plot(COVS, treat = test.data$cohort, weights =
7 test.data$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
8 var.order = "adjusted", title = "B1 - BARI vs TNFi", color =
9 c("#FD8D3C", "#08306B"), themes = theme_pubclean() )
10
11 # BARI vs OMA
12 B2 <- love.plot(COVS_2, treat = test.data2$cohort, weights =
13 test.data2$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
14 var.order = "adjusted", title = "B2 - BARI vs OMA", color =
15 c("#FD8D3C", "#08306B") , themes = theme_pubclean() )
16
17 one <- ( pscore_plot + B1)
18
19 two <- ( pscore_plot2 + B2 )
20
21 png("./3_clean_output/figures/AIPTW diagnostc COLOR.png", width =
22 1300, height = 900, res = 100) # opening graphic device
23
24 one / two
25
26 dev.off() # closing graphic device
27
28
29
30
```

1. [3] Fist line analysis

Non-adjusted Survival curves

BARI vs TNFi

```
31
32
33 BARI_first1 <- copy(BARI_first[cohort %in% c("BARI", "TNFi")])
34
35
36
37
38
39
40 surv_object3 <- Surv(time = BARI_first1$time_on_drug, event =
41 BARI_first1$stop_DMARD) # indiqate stop variable and time_on_drug
42 summary(coxph(surv_object3 ~ cohort, data = BARI_first1))
43 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first1) # function
44 which creates Kaplan-meier data
45 survplot_first1 <- ggsurvplot(fit3, data = BARI_first1, # plot
46 pval = T,
47 pval.method = TRUE,
48 legend.title = "Groups :",
49 legend.labs = c("BARI", "TFNi"),
50 xlab = "Time (days)",
51 xlim = c(0, 700),
52 censor = FALSE,
53 title = "A - BARI vs TNFi",
54 surv.median.line = "v",
55 linetype = 1,
56
57
58
59
```

```

1
2
3     size = 1.5,
4     ggtheme = theme_benoit(),
5     # palette = c("grey78", "grey50"),
6     palette = c("red2", "green3"), # to get colors
7     risk.table = T
8   )
9
10  survplot_first1$plot$labels$y <- "Proportion still on drug" # to
11  change the label
12  survplot_first1
13
14
15  rm(surv_object3, fit3)
16
17  saving plot curves
18
19  png("./3_clean_output/PLOT BARI vs TNFi first line curves non adjusted
20  COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
21  device
22
23  survplot_first1
24
25  dev.off() # closing graphic device
26
27  BARI vs OMA
28
29  BARI_first2 <- BARI_first[line_of_therapy == "1st" & cohort %in%
30  c("BARI", "OMA")] # selection des TC TNFi
31
32
33  surv_object3 <- Surv(time = BARI_first2$time_on_drug, event =
34  BARI_first2$stop_DMARD) # indiquer stop variable and time_on_drug
35  summary(coxph(surv_object3 ~ cohort, data = BARI_first2))
36  fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first2) # fonction
37  which creates Kaplan-meier data
38  survplot_first2 <- ggsurvplot(fit3, data = BARI_first2, # plot
39    pval = T,
40    pval.method = TRUE,
41    legend.title = "Groups :",
42    legend.labs = c("BARI", "OMA"),
43    xlab = "Time (days)",
44    xlim = c(0, 700),
45    censor = FALSE,
46    title = "B - BARI vs OMA",
47    surv.median.line = "v",
48    linetype = 1,
49    size = 1.5,
50    ggtheme = theme_benoit(),
51    # palette = c("grey78", "grey50", "grey10"),
52    palette = c("red2", "blue3"), # to get colors
53    risk.table = T
54  )
55
56
57
58
59
60

```

```
1
2
3
4  survplot_first2$plot$labels$y <- "Proportion still on drug" # to
5  change the label
6  survplot_first2
7
8
9  rm(surv_object3, fit3)
10
11 saving plot curves
12
13 png("./3_clean_output/PLOT BARI vs OMA first line curves non adjusted
14 COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
15 device
16
17 survplot_first2
18
19 dev.off() # closing graphic device
20
21 Adjusted curves with imputed data (BARI vs TNFi)
22 dummy_cox_impute_first1 <- mice::complete(imputed_data1_first, "long",
23 include = T)
24 dummy_cox_impute_first1 <-
25 dummy_cox_impute_first1[dummy_cox_impute_first1$.imp != 0,]
26
27 BARI_first1_fit <- survfit(coxph(Surv(time = time_on_drug, event =
28 stop_DMARD) ~ cohort+
29
30     I(age_base/10)+
31     bmi_base+
32     concomitant_csDMARD+
33     PREDNISON_STEROID+
34     CDAI0+
35     I(disease_duration_base_years/10)+
36     C(smoker_base, base=3)+
37     line_of_therapy+
38     gender+
39     seropositivity_base+
40     cluster(patient_id)+
41     strata(cohort), dummy_cox_impute_first1),
42 data = dummy_cox_impute_first1)
43
44
45
46 survplot_first1_adj <- ggsurvplot(BARI_first1_fit, data =
47 dummy_cox_impute_first1, variable = "cohort",
48     xlab = "Time (days)",
49     title = "Multivariable Cox model of drug discontinuation by
50 type of treatment - 1st line vs 1st line",
51     legend.title = "Groups :",
52     legend.labs = c("Baricitinib", "TNFi"),
53     censor = FALSE,
54     xlim = c(0, 700),
55     surv.median.line = "v",
56
57
58
59
```

```

1
2
3     linetype = 1,
4     size = 1.5,
5     ggtheme = theme_minimal(),
6     #palette = c("grey78", "grey10")
7     palette = c("red2", "green3"), # to get colors
8   )
9
10  survplot_first1_adj <- survplot_first1_adj +
11    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
12    concomitant glucocorticoid, baseline CDAI, disease duration (decades),
13    smoking status, line of therapy, gender, seropositivity")
14
15
16  survplot_first1_adj
17  table(BARI_first1$cohort)
18  rm(dummy_cox_impute_first1, BARI_first1_fit)
19
20  Saving the survival plot in PNG file
21
22  png("./3_clean_output/PLOT BARI vs TNFi first line curves adjusted
23  COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
24  device
25
26  survplot_first1_adj
27
28  dev.off() # closing graphic device
29
30  Adjusted curves with imputed data (BARI vs OMA)
31  dummy_cox_impute_first2 <- mice::complete(imputed_data2_first, "long",
32  include = T)
33  dummy_cox_impute_first2 <-
34  dummy_cox_impute_first2[dummy_cox_impute_first2$.imp != 0,]
35
36
37  BARI_first2_fit <- survfit(coxph(Surv(time = time_on_drug, event =
38  stop_DMARD) ~ cohort+
39    I(age_base/10)+
40    bmi_base+
41    concomitant_csDMARD+
42    PREDNISON_STEROID+
43    CDAI0+
44    I(disease_duration_base_years/10)+
45    C(smoker_base, base=3)+
46    line_of_therapy+
47    gender+
48    seropositivity_base+
49    cluster(patient_id)+
50    strata(cohort), dummy_cox_impute_first2),
51  data = dummy_cox_impute_first2)
52
53
54
55  survplot_first2_adj <- ggsurvplot(BARI_first2_fit, data =
56
57
58
59
60

```

```
1
2
3 dummy_cox_impute_first2, variable = "cohort",
4     xlab = "Time (days)",
5     title = "Multivariable Cox model of drug discontinuation by
6 type of treatment - 1st line vs 1st line",
7     legend.title = "Groups :",
8     legend.labs = c("Baricitinib", "OMA bDMARDs"),
9     censor = FALSE,
10    xlim = c(0, 700),
11    surv.median.line = "v",
12    linetype = 1,
13    size = 1.5,
14    ggtheme = theme_minimal(),
15    #palette = c("grey78", "grey50")
16    palette = c("red2", "blue3"), # to get colors
17  )
18
19
```

```
20 survplot_first2_adj <- survplot_first2_adj +
21     labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
22 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
23 smoking status, line of therapy, gender, seropositivity")
24
```

```
25
26 survplot_first2_adj
27 table(BARI_first2$cohort)
28 rm(dummy_cox_impute_first2, BARI_first2_fit)
29
```

30 Saving the survival plot in PNG file

```
31
32 png("./3_clean_output/PLOT BARI vs OMA first line curves adjusted
33 COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
34 device
35
```

```
36 survplot_first2_adj
37
```

```
38 dev.off() # closing graphic device
39
```

40 Multipanel plots

41 Non adjusted curves

```
42
43 plots <- list()
44 plots[[1]] <- survplot_first1
45 plots[[2]] <- survplot_first2
46
47
```

48 # Nice function

```
49
50 multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
51
```

```
52
53 png("./3_clean_output/figures/BILOT BARI vs TNFi vs OMA 1st Line
54 curves non-adjusted COLOR.png", width = 1000, height = 600, res = 100)
55 # opening graphic device
56
57
58
59
```

```
1
2
3
4 multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
5
6 dev.off() # closing graphic device
7
8 All in one BARI vs TNFi vs OMA
9
10 BARI_first[,time_on_drug_year := time_on_drug/365.25]
11
12 surv_object4 <- Surv(time = BARI_first$time_on_drug_year, event =
13 BARI_first$stop_DMARD)
14 fit4 <- survfit(surv_object4 ~ cohort, data = BARI_first) # this
15 function creates the data for Kaplan Meyer
16
17
18 survplot_4 <- ggsurvplot(fit4, data = BARI_first, # plot
19 pval = F,
20 pval.method = F,
21 legend.title = "Groups",
22 legend.labs = c("BARI", "TNFi", "OMA"),
23 xlab = "Time (years)",
24 xlim = c(0, 2.5),
25 censor = FALSE,
26 # title = "Non-adjusted drug discontinuation by type of
27 treatment (Kaplan-Meier)",
28 surv.median.line = "v",
29 linetype = 1,
30 size = 1.5,
31 ggtheme = theme_benoit(),
32 palette = c("red3", "green2", "blue2"), # to put colors
33 risk.table = T)
34
35
36 values <- summary(fit4)$table[, "median"]
37 df <- data.frame(y = .2, x = values+.2, label =
38 as.character(paste(round(values*365.25, 2), "\n days")))
39 df <- df[2,]
40
41 survplot_4$plot <- survplot_4$plot +
42 geom_text(data = df, aes(x, y, label = label), color = c("green2"),
43 size = 5)
44
45
46 survplot_4$plot$labels$y <- "Proportion still on drug" # to change the
47 label
48 survplot_4
49
50 Saving survplot
51
52 png("./3_clean_output/figures/PLOT BARI vs TNFi vs OMA first curves
53 non adjusted COLOR.png", width = 800, height = 600, res = 100) #
54 opening graphic device
55
56
57
58
59
60
```

```

1
2
3 survplot_4
4
5 dev.off() # closing graphic device
6
7

```

1. [4] LACK of EFFICACY and ADVERSE EVENTS

Analysis by stop_reasons in competing risk

(BARI vs TNFi)

Cumulative incidence function

```

16 ci_long$time_months <- ci_long$time/365.25*12
17
18
19 plot2 <- ggplot(data = ci_long, aes(x = time_months,
20                                     y = value,
21                                     linetype = variable ,
22                                     col = cohort )) +
23
24   geom_line(size = 0.75)+
25
26   scale_color_manual(breaks = c(0,1),
27                       values = c("red3","green2"),
28                       labels = c("BARI","TNFi"))+
29   scale_linetype_manual(breaks = c("CI.1","CI.2"),
30                          values = c("solid","dashed", "dotted"),
31                          labels = c("Adverse Event","Ineffectiveness"))
32 + # not showing the "other category"
33   scale_x_continuous(name = "Time (months)",
34                      breaks = c(0,3,6,9,12),
35                      limits = c(0,12)) +
36   scale_y_continuous(name = "Cumulative incidence", limits = c(0,0.4))
37 +
38   theme_benoit()+
39   theme(legend.box = "horizontal",
40         legend.position = c(0.05,1),
41         legend.justification = c(0,1),
42         legend.background = element_blank(),
43         legend.key = element_blank(), #no legend key background
44         legend.key.width = grid::unit(2, "lines"))+ # longer line in
45 legend, to see properly the dashed
46   labs(color = "", linetype = "", title = "A - BARI vs TNFi")
47
48
49 plot2
50
51 ggsave(filename = "PLOT BARI vs TNFi cumulative incidence.png",plot =
52 plot2, path = "./3_clean_output/", device = "png", width = 829, height
53 = 550, units = "px", scale = 3.2)
54
55
56
57
58
59
60

```

(BARI vs OMA)*Cumulative incidence function*

```

1
2
3
4
5
6 ci_long_2$time_months <- ci_long_2$time/365.25*12
7
8
9 plot3 <- ggplot(data = ci_long_2, aes(x = time_months,
10                                     y = value,
11                                     linetype = variable ,
12                                     col = cohort )) +
13
14   geom_line(size = 0.75)+
15
16   scale_color_manual(breaks = c(0,1),
17                     values = c("red3","blue2"),
18                     labels = c("BARI","OMA"))+
19   scale_linetype_manual(breaks = c("CI.1","CI.2"), # not showing the
20 "other" category
21                       values = c("solid","dashed", "dotted"),
22                       labels = c("Adverse Event","Ineffectiveness"))
23
24 +
25   scale_x_continuous(name = "Time (months)",
26                     breaks = c(0,3,6,9,12),
27                     limits = c(0,12)) +
28   scale_y_continuous(name = "Cumulative incidence", limits = c(0,0.4))
29
30 +
31   theme_benoit()+
32   theme(legend.box = "horizontal",
33         legend.position = c(0.05,1),
34         legend.justification = c(0,1),
35         legend.background = element_blank(),
36         legend.key = element_blank(), #no legend key background
37         legend.key.width = grid::unit(2, "lines"))+ # longer line in
38 legend, to see properly the dashed
39   labs(color = "", linetype = "", title = "B - BARI vs OMA")
40
41 plot3
42
43 ggsave(filename = "./3_clean_output/figures/PLOT BARI vs OMA
44 cumulative incidence.png", plot3, height = 4, width = 6, units =
45 "in",dpi = 300)

```

Multipanel

```

46
47 plot2_3 <- plot2 + plot3
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

59 ggsave(filename = "./3_clean_output/figures/PLOT BARI vs TNFi and BARI
60 vs OMA cumulative incidence.png", plot2_3, height = 4, width = 8,
61 units = "in",dpi = 300)

```


1. [6] LDA - REM

Exploration

See all available raw CDAI measures : :)

```

BARI_long[, group := "non-BARI"]
BARI_long[drug == "BIOLOGIC_BARICITINIB", group := "BARI"]

plot_data <- copy(BARI_long[!is.na(TC_id) & TC_id %in%
BARI_DATA$TC_id])
plot_data <- merge(plot_data, BARI_DATA[,.(TC_id, cohort)], by =
"TC_id")

CDAI_plot <- ggplot(data = plot_data,
                    aes(x = time, y = CDAI, fill = cohort) )+

  annotate("rect",xmin = 0.875,xmax = 1.125,
          ymin = -1,ymax = 60,alpha = .5,fill = "grey80")+
  annotate("rect",xmin = -.05,xmax = 0+1.5/12,
          ymin = -1,ymax = 60,alpha = .5,fill = "grey80")+

  geom_point(data = plot_data[cohort != "BARI"], alpha = 0.2, size =
2, shape = 21, position = position_jitter(width = 0.02, seed = 123) )+
  geom_point(data = plot_data[cohort == "BARI"], alpha = 0.25, size =
2, shape = 21, position = position_jitter(width = 0.02, seed = 123) )+

  #geom_jitter(width = 0.01, height = 0.01, data = plot_data[cohort !=
"BARI"], alpha = 0.2, size = 2, shape = 21, show.legend = F )+
  #geom_jitter(width = 0.01, height = 0.01, data = plot_data[cohort ==
"BARI"], alpha = 0.25, size = 2, shape = 21 , show.legend = F, )+

  geom_smooth(alpha = 0.1, size = 1, aes(color = cohort), show.legend
= F)+

  coord_cartesian(xlim = c(0,2.5))+
  labs(title = "CDAI across time type of treatment (all TC)",
       x = "Time (years since TC initiation)",
       y = "CDAI score",
       color = "",
       fill = "")+
  theme_benoit()+
  theme(legend.position = c(1,1),
        legend.justification = c(1,1))+
  guides(color = guide_legend(override.aes = list(linetype = NA,size =
3)))

CDAI_plot

```

To me this figure is the best results to be discussed regarding REM and LDA

saving plot

```
png("./3_clean_output/figures/PLOT CDAI across time raw.png",  
     width = 8,  
     height = 6,  
     units = "in",  
     res = 120) # opening graphic device
```

CDAI_plot

```
dev.off() # closing graphic device
```

CARRAC histogram

Building large format data table from the CARRAC output

```
# Extracting LDA BARI  
LDA_BARI <- rbind(LDA_BARI_TNF[2,1:4], LDA_BARI_OMA[2,1:4]) # I have  
one estimation per comparison  
LDA_BARI[, LDA := mean(LDA)][, LDA_sup := mean(LDA_sup)][, LDA_inf :=  
mean(LDA_inf)] # averaging  
LDA_BARI <- LDA_BARI[1]
```

```
# REM BARI  
REM_BARI <- rbind(REM_BARI_TNF[2,1:4], REM_BARI_OMA[2,1:4]) # I have  
one estimation per comparison  
REM_BARI[, LDA := mean(LDA)][, LDA_sup := mean(LDA_sup)][, LDA_inf :=  
mean(LDA_inf)] # averaging  
REM_BARI <- REM_BARI[1]
```

```
# IDem for TNFi and OMA  
LDA_TNFi <- LDA_BARI_TNF[3,1:4]  
REM_TNFi <- REM_BARI_TNF[3,1:4]
```

```
LDA_OMA <- LDA_BARI_OMA[3,1:4]  
REM_OMA <- REM_BARI_OMA[3,1:4]
```

```
# Binding together  
LDA <- rbind(LDA_BARI, LDA_TNFi, LDA_OMA)  
setnames(LDA, c("ttt", "LDA", "LDA_sup", "LDA_inf")) #putting right  
labels  
REM <- rbind(REM_BARI, REM_TNFi, REM_OMA)  
setnames(REM, c("ttt", "REM", "REM_sup", "REM_inf"))
```

```
histo_carrac <- cbind(LDA, REM[, -1])
```

plotting

```
1
2
3 carrac_plot <- ggplot(data = histo_carrac, aes(x = ttt, group = ttt))
4 +
5
6   theme_pubclean()+
7
8   geom_errorbar( mapping=aes(x=ttt, ymin=LDA_inf*100,
9 ymax=LDA_sup*100), width=0.2, size=1, color="grey70")+
10   geom_errorbar( mapping=aes(x=ttt, ymin=REM_inf*100,
11 ymax=REM_sup*100), width=0.2, size=1, color="grey50")+
12
13   geom_bar(aes(y = LDA*100), stat = "identity", fill = "grey80", alpha
14 = 0.5)+
15   geom_text(aes(y = LDA*100, label = "LDA"), vjust = 1.5) +
16
17   geom_bar(aes(x = ttt, y = REM*100), stat = "identity", fill =
18 "grey65", alpha = 0.5)+
19   geom_text(aes(x= ttt, y = REM*100, label = "REM"), vjust=1.5) +
20
21   theme(strip.text.y = element_blank(),
22         strip.background = element_blank(),
23         axis.line.x = element_line(size = 0.5),
24         axis.text = element_text(face = "bold", colour = "black"),
25         legend.position="bottom", plot.margin =
26         unit(c(1,3,2,1),"lines"))+
27
28   scale_y_continuous(limits = c(0,82))+
29
30   labs(y = "(% of TC)", x = "Treatment group", title = "A - REM and
31 LDA rates \nby type of treatment \n(CARRAC)")
32
33 carrac_plot
34
35 also Saving CARRAC plot only
36
37 png("./3_clean_output/figures/PLOT BARI 3 CARRAC ONLY.png", width =
38 350, height = 600, res = 100) # opening graphic device
39
40 ggplot(data = histo_carrac, aes(x = ttt, group = ttt)) +
41
42   theme_pubclean()+
43
44   geom_errorbar( mapping=aes(x=ttt, ymin=LDA_inf*100,
45 ymax=LDA_sup*100), width=0.2, size=1, color="grey70")+
46   geom_errorbar( mapping=aes(x=ttt, ymin=REM_inf*100,
47 ymax=REM_sup*100), width=0.2, size=1, color="grey50")+
48
49   geom_bar(aes(y = LDA*100), stat = "identity", fill = "grey80", alpha
50 = 0.5)+
51   geom_text(aes(y = LDA*100, label = "LDA"), vjust = 1.5) +
52
53
54
55
56
57
58
59
60
```

```
1
2
3     geom_bar(aes(x = ttt, y = REM*100), stat = "identity", fill =
4 "grey65", alpha = 0.5)+
5     geom_text(aes(x= ttt, y = REM*100, label = "REM"), vjust=1.5) +
6
7     theme(strip.text.y = element_blank(),
8           strip.background = element_blank(),
9           axis.line.x = element_line(size = 0.5),
10          axis.text = element_text(face = "bold", colour = "black"),
11          legend.position="bottom", plot.margin =
12 unit(c(1,3,2,1),"lines"))+
13
14     scale_y_continuous(limits = c(0,82))+
15
16     labs(y = "(% of TC)", x = "Treatment group", title = "REM and LDA
17 rates \nby type of treatment \n(CARRAC)")
18
19
20 dev.off() # closing graphic device
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	p2 p2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	p4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	p4-5
Methods			
Study design	4	Present key elements of study design early in the paper	p6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	p6-7
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	p6 p6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	p7 & supp p2
Data sources/measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	supp p2
Bias	9	Describe any efforts to address potential sources of bias	p16
Study size	10	Explain how the study size was arrived at	supp p9
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	p7-8
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	p8-9 supp p 5-6 p8-9 p8-9 p9 & supp p5-6

Continued on next page

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60**Results**

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	p10
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	p11-12
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	p13
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	p13-14
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	p14-15 & supp

Discussion

Key results	18	Summarise key results with reference to study objectives	p15
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	p16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	p17
Generalisability	21	Discuss the generalisability (external validity) of the study results	p17
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	p18

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

Filled in by Benoît GILBERT, 30-01-2023

BMJ Open

COMPARATIVE EFFECTIVENESS OF BARICITINIB AND ALTERNATIVE BIOLOGICAL DMARDs IN A SWISS COHORT STUDY OF RA PATIENTS

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2023-072300.R2
Article Type:	Original research
Date Submitted by the Author:	09-Feb-2024
Complete List of Authors:	Gilbert, Benoit Thomas P.; Geneva University Hospitals, Division of Rheumatology, Department of Internal Medicine Specialties Mongin, Denis; Hopitaux Universitaires de Geneve, Rheumatology, Department of Internal Medicine Specialties Aymon, Romain; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties Lauper, Kim; Hopitaux Universitaires de Geneve, Rheumatology, Department of Internal Medicine Specialties LAEDERMANN, Cédric; Eli Lilly PERRIER, Clémentine; Eli Lilly Mueller, Ruediger; Rheumazentrum Ostschweiz, private practice Courvoisier, Delphine; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties Finckh, Axel; Geneva University Hospitals, Rheumatology, Department of Internal Medicine Specialties
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Epidemiology
Keywords:	RHEUMATOLOGY, EPIDEMIOLOGY, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

COMPARATIVE EFFECTIVENESS OF BARICITINIB AND ALTERNATIVE BIOLOGICAL DMARDs IN A SWISS COHORT STUDY OF RA PATIENTS

Benoit GILBERT^{1*}, Denis MONGIN¹, Romain AYMONT¹, Kim LAUPER^{1,2}, Cédric LAEDERMANN³, Clémentine PERRIER³, Ruediger MUELLER⁴, Delphine S. COURVOISIER¹, Axel FINCKH¹

¹ Division of Rheumatology, Geneva University Hospitals, Geneva.

² Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, University of Manchester.

³ Eli Lilly (Suisse) SA, 16 Ch Des Coquelicots, CH-1214 Vernier.

⁴ Private practice, St Gallen, Switzerland.

* Correspondence:

Benoît Thomas P. GILBERT

26 Avenue de Beau-Séjour

Division of Rheumatology, Department of Medicine, Geneva University Hospital,

1206 Geneva,

Switzerland

benoit.gilbert@unige.ch

ABSTRACT

Objectives: This observational study compares the effectiveness of baricitinib (BARI), a targeted synthetic DMARD (tsDMARD) with alternative biological DMARDs (bDMARDs) in rheumatoid arthritis (RA) patients, from a prospective, longitudinal cohort.

Methods: We compared patients initiating a treatment course of BARI, tumor necrosis factor inhibitors (TNFi) or other mode of action bDMARDs (OMA), during a period when all these DMARDs were available in Switzerland. The primary outcome was drug-maintenance; secondary outcomes included discontinuation rates related specifically to ineffectiveness and to adverse events. We further analyzed rates of low disease activity (LDA) and remission (REM) at 12 months, and drug maintenance in b- and tsDMARD-naïve population.

Results: A total of 1053 treatment courses (TC) were included: 273 on BARI, 473 on TNFi and 307 on OMA. BARI was prescribed to older patients with longer disease duration and more previous treatment failures than TNFi. Compared to BARI, the adjusted drug maintenance was significantly shorter for TNFi (hazard ratio (HR) for discontinuation: 1.76; 95% CI [1.32-2.35]), but not compared to OMA (HR 1.27; 95% CI [0.93-1.72]). These results were similar in the b/tsDMARD-naïve population. The higher discontinuation of TNFi was mostly due to an increased discontinuation for ineffectiveness (HR = 1.49; 95% CI [1.03 – 2.15]), with no significant differences in drug discontinuation for adverse events (HR = 1.46; 95% CI [0.83 - 2.57]). The LDA and REM rates at 12 months did not differ significantly between the 3 groups.

Conclusions: BARI demonstrated a significantly higher drug maintenance compared to TNFi, mainly due to lower drug discontinuations for ineffectiveness. We found no difference in drug-maintenance between BARI and OMA. Clinical outcomes did not differ between the three groups. Our results suggest that BARI is an appropriate therapeutic alternative to bDMARDs in the management of RA.

Strengths and limitations of this study

Strengths:

- Use data derived from office-based rheumatologists
- Study period where all alternative medications were available on the market
- Several sensitivity analyses, congruent with main results

Limitations:

- Not a randomized setting
- Sub-analysis in b/tsDMARD-naïve population has limited sample size

INTRODUCTION

Rheumatoid arthritis (RA) is an auto-immune disease leading to widespread inflammation and irreversible joint damage, if insufficiently treated. New treatment paradigms have emerged in the last decades, such as “early aggressive therapy” in the so called “window of opportunity”, during which patients are more likely to reach long term remission.[1] A wide panel of biological disease modifying antirheumatic drugs (bDMARDs) and targeted synthetic DMARDs (tsDMARDs) have been approved in the management of RA, after failure of methotrexate. In clinical-trial settings, b- and tsDMARDs have demonstrated significant reduction of joint inflammation and prevention of joint damage.[2–8]

Efficacy estimates from placebo-controlled randomized trials often differ from real-world effectiveness estimates, because of patient selection, adherence to therapy and other reasons.[9–12] Indeed, drug maintenance of many bDMARDs remains modest in observational analyses, while long term remissions are rare and secondary loss of efficacy frequent.[13] Furthermore, understanding the clinical effectiveness of bDMARDs or tsDMARDs in specific conditions, such as elderly or multi-morbid patients, may become important as we move towards personalized care. Finally, trials provide only limited data on long term effectiveness and safety because clinical-trial follow-up is typically less than 12 months.

Baricitinib (BARI) has been approved in Switzerland for the treatment of RA in 2017 as well as all around the world. Clinical trials with BARI have established efficacy and demonstrated acceptable adverse events profile, both in combination with methotrexate or in monotherapy.[14–20] However, evidence about effectiveness of BARI compared to TNFi in

1
2
3 real-world settings are scarce. A recently published analysis of registry data from Sweden
4
5 showed that baricitinib had higher maintenance as compared to most other bDMARD.[21]
6
7 Pappas et al., in the United States, also demonstrated that TNFi and non-TNFi drugs had
8
9 similar outcomes when prescribed in b/tsDMARD-naïve population, an observation replicated
10
11 in the RA-BE-REAL study.[22,23]
12
13

14
15 The aim of our analysis was to compare real-world drug maintenance between BARI and other
16
17 approved b/tsDMARDs, using data from a European registry.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODS

Study population

This is a nested cohort study from a prospective, longitudinal, cohort of Swiss RA patients in a real-life setting, the Swiss Clinical Quality Management registry (SCQM). The SCQM registry was founded in 1997 with the financial support of Swiss regulatory authorities, who recommended a continuous monitoring of all patients receiving new DMARDs. Unlike many other European registries, most patients are enrolled by private office-based rheumatologists (60%), providing a population-based sample of RA patients in Switzerland. All approved RA treatments are represented in the registry. The data for this analysis was extracted from the SCQM registry on 2020-06-01.

We used “treatment courses” (TCs) as our denominator of interest, with each new treatment initiation considered as a separate “treatment course” (TC). We included all TCs with the medications of interest initiated between 2017-09-01 and 2020-06-01, with at least one follow-up visit, in adult patients with a diagnosis of RA confirmed by a rheumatologist. Thus, a given patient could potentially contribute to several TCs during the study period. To minimize the risk of confounding bias, the time window was selected to include only the period when all the therapies examined were available for prescription and reimbursed (BARI was first reimbursed on the Swiss market in September 2017). We excluded TCs with no follow-up visit at the time of data extraction.

Exposure of interest

The exposure of interest was the type of treatment used, namely BARI, TNFi, and other mode of action bDMARDs (OMA), excluding other tsDMARDs and rituximab. We decided to exclude rituximab a priori, because its long-term action impairs precise estimation of treatment discontinuation. Tofacitinib was excluded because we had insufficient TC to perform meaningful comparative effectiveness analyses against a single other specific tsDMARD agent. Included TNFi treatments were: adalimumab, etanercept, golimumab, certolizumab, infliximab. Included OMA treatments were: tocilizumab, abatacept, sarilumab, and anakinra.

Outcomes

The primary outcome of this analysis was the time to all-cause-discontinuation. This outcome, also referred to as “drug maintenance”, captures both the drug’s effectiveness and its tolerance.[24] The time to all-cause-discontinuation was defined as the number of days between treatment initiation and the reported date of discontinuation, or the date of initiation of a new b/tsDMARD, whatever came first. In survival analyses, death or lost-to-follow up are censored. We also report discontinuation rates at 12 months. Temporary discontinuations of less than 6 months (for instance, because of an elective surgery or a pregnancy) were not considered a permanent drug discontinuation. Discontinuation reasons are recorded by the clinician when stopping a DMARD treatment, who chooses between four options (“Adverse event”, “Ineffectiveness”, “Remission”, or “Other”).

Pre-planned secondary outcomes, were time to discontinuation due to ineffectiveness and time to discontinuation due to adverse events. Other secondary outcomes included response rates, namely the rates of low disease activity (LDA) and remission (REM), at 12 months, defined respectively as attaining a CDAI score ≤ 10 and CDAI score ≤ 2.8 (not mutually

1
2
3 exclusive).[25] Finally, we performed an exploratory subgroup analysis, restricting the
4
5 population to b/tsDMARD-naive patients only, and re-assessing the main outcome in this
6
7 setting.
8
9

10 **Statistical analysis**

11
12
13 Analyses were conducted and reported in accordance to EULAR recommendations for
14
15 comparative effectiveness research.[9] Baseline characteristics were compared using
16
17 generalized linear mixed models to account for repeated treatments within the same patients.
18
19 For the primary outcomes, Kaplan-Meier survival analyses were used to assess crude drug
20
21 maintenance, and groups were compared using Log-Rank tests. Subsequently, missing
22
23 covariates were imputed using chained equations (see below for details). We then
24
25 implemented Cox proportional hazard ratio models, to obtain adjusted estimates. Based on
26
27 prior subject matter knowledge,[26] we adjusted our models for the following potential
28
29 confounders: age, gender, BMI, concomitant csDMARDs use (yes/no), concomitant
30
31 prednisone usage (yes/no), CDAI score at baseline, disease duration, smoking status (current-,
32
33 former-, never-smoker) and line of therapy (1st, 2nd, 3rd, 4th or more), and seropositivity
34
35 (yes/no). Detailed definitions for each variable are available in the supplement
36
37 (Supplementary Material 1). The main analysis (survival analysis) accounted for clustering
38
39 resulting from patients with multiple treatment courses, inducing correlation within the
40
41 patient-level data. The cluster term is used to compute a robust variance for the model, by
42
43 applying the so-called Huber sandwich estimator.[27] . All conditions of application of the Cox
44
45 model were verified. One additional sensitivity analysis was conducted for the primary
46
47 outcome, using augmented inverse probability of treatment weighting (AIPTW).
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 In secondary analyses, we used the Fine-Gray approach to assess specific reasons for drug
4 discontinuation (i.e. ineffectiveness, or adverse event) in a competing-risk setting. The Fine-
5 Gray method takes competing risks into account when estimating the cumulative incidence
6 function, modelling the sub-distribution hazard without treating competing events as
7 censoring events.[28] Other secondary outcomes included response rates (LDA and REM) at
8 12 months. To avoid overestimations, we computed the response rates using the ‘confounder-
9 adjusted response rate with attrition correction’ (CARRAC) method.[29] The latter estimates
10 the response rates using multiple imputations, with a model including both confounders and
11 treatment stop reason. CARRAC thus provides reliable estimates when reasons for treatment
12 discontinuation differ between compared groups.

13
14 For all adjusted analyses, missing baseline covariates were imputed using closest value in a
15 window of -90 to +30 days. However, this window was reduced to -30 to +7 days when
16 imputing baseline CDAI. If still missing after this first step, baseline CDAI values were imputed
17 using linear mixed effect regression model with quadratic time. We imputed other baseline
18 covariates with chained equations technique, which provides unbiased estimates if the
19 variables are missing at random.[29] Such imputations were performed using 50 datasets with
20 25 iterations. Imputation was done using the whole data set, before adequately subsetting
21 the data for each group comparison.

22
23 We also imputed data required for secondary outcomes, including disease activity. If the CDAI
24 score at 12-month was not available, the closest value in a window of +/- 45 days was used (a
25 3-month-wide window). If still missing, the 12-month CDAI values were imputed using nearest
26 neighboring value, as previously described.[30]

27
28 All analyses were conducted using R (version 4.0.3), in particular with packages “*tableone*”,
29 “*survival*”, “*mice*”.[31] Two tailed p-values < 0.05 were considered significant. We did not
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 adjust p value for multiple comparisons, as outcomes were pre-specified. Final analysis code
4
5 is shown in the supplement (Supplementary Material 2).
6
7

8 **Patient and Public Involvement**

9

10 Patient involvement is central to the SCQM cohort. Several patients are part of the executive
11
12 board and involved in the approval of research projects.
13
14
15
16
17

18 **RESULTS**

19

20 **Population description**

21

22
23 During the study period, 1053 TC were initiated in 834 different patients, including 273 TC
24
25 with BARI, 473 with TNFi and 307 with OMA (Figure S1; Figure S2). TNFi were more often given
26
27 as a second line therapy after methotrexate failure. Inversely, BARI was prescribed to
28
29 significantly older patients, with longer disease durations and more previous treatment
30
31 failures (Table 1).
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Baseline characteristics of study population, SCQM-RA registry, 2017-2020.

Variable	BARI (TC = 273; 273 pts)		TNFi (TC = 473; 408 pts)		OMA (TC = 307; 289 pts)		p values
	n % of total in group Otherwise: Mean (SD)						
Patients		Miss.		Miss.		Miss.	
Female	78 %	0	74 %	1	73 %	1	0.097
Age (years)	59 (14)	0	52 (15)	1	59 (13)	1	0.021
Disease duration (years)	13 (10)	4	8 (9)	19	11 (9)	5	0.027
CDAI baseline (raw data)	19 (10)	175	18 (10)	301	20 (13)	204	0.34
CDAI baseline (imputed)	15 (9)	0	14 (9)	0	16 (11)	0	0.06
Obesity (BMI > 30)	16 %	104	14 %	134	13 %	115	0.85
Smoking	17 %		18 %		21 %		Ref.
Current	28 %	32	26 %	69	28 %	26	0.95
Former	43 %		41 %		48 %		0.98
Never							0.92
Seropositive (ACPA or RF)	75 %	1	70 %	7	77 %	5	
TC		Miss.		Miss.		Miss.	
Concomitant csDMARD	40 %	0	61 %	0	46 %	0	<0.01
Line of Therapy	17 %		48 %		22 %		Ref.
1 st (= bio-naïve)	20 %	0	23 %	0	24 %	0	<0.01
2 nd	19 %		11 %		24 %		<0.01
3 rd	44 %		18 %		31 %		<0.01
4 th or later							<0.01
Previous tsDMARD use (non-BARI)	33 %	0	1 %	0	5 %	0	<0.01
Concomitant glucocorticoid (at any time)	22 %	0	20 %	0	24 %	0	0.35
Mean dose of concomitant glucocorticoid (mg)	2.0 (4.6)	0	2.1 (5.4)	0	2.2 (5.1)	0	0.95
Dose of BARI (4mg)	86 %	0	-	-	-	-	-

Table 1 Legend: In Switzerland, BARI was prescribed to older patients, with longer disease duration and more previous treatment failures. Missing values for covariables are reported as absolute numbers

BARI = baricitinib. TC = Treatment Courses. CDAI = Clinical Disease Activity Index. TNFi = Tumor Necrosis Factor Inhibitors. OMA = bDMARDs with Other Mode of Action. tsDMARD = targeted synthetic DMARDs. ACPA = Anti Citrullinated Peptide Antibodies. RF = Rheumatoid Factor. Miss. = number of missing values. Ref. = Reference

1
2
3 category for p-values (pairwise comparisons). P-values are obtained by generalized linear mixed models to
4 account for repeated treatments within the same patients. In TFNi and OMA groups, some patients have
5 contributed several TC, thus total number of TCs exceeds total number of patients.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Time to all-cause-discontinuation

Crude proportions of treatment discontinuation by reasons are reported in Table S1, and crude times of observation are represented on Figure S2.

At 12 months, based on the Kaplan-Meier curves (Figure 1), the estimated proportions of patients still on therapy were: 71% (95% CI [65% - 77%]) in the BARI group, 55% (95% CI [50% - 61%]) in the TNFi group, and 63% (95% CI [57% - 70%]) in the OMA group.

Overall, unadjusted time to all-cause-discontinuation was significantly longer in the BARI group compared to the TNFi group (estimated median prescription survival-time of 704 versus 448 days; Log-rank $p < 0.01$; Figure 1). These results persisted after adjustment for confounding factors using the multivariable Cox model (HR = 1.76; 95% CI [1.32-2.35]; $p < 0.001$; Table S2 and Figure S3; Figure S4).

BARI versus OMA time to all-cause-discontinuation was not significantly different, even after adjustment (HR 1.27; 95% CI [0.93-1.72]; $p = 0.13$; Table S2, Figure S3 and Figure S4).

Sensitivity analyses using AIPTW led to similar conclusions (Figure S5). Covariates significantly associated with decreased drug maintenance were high baseline CDAI scores and concomitant glucocorticoid usage (Table S2 and Figure S3).

Time to all-cause-discontinuation in b/tsDMARD-naïve patients

In this exploratory subgroup analysis, we restricted the population to patients without prior experience of b/tsDMARDs (so-called 'bio-naïve' patients, i.e. first b/tsDMARD prescription after methotrexate failure). In this subpopulation, patient characteristics were more balanced than in the main analysis, except for age, which remained younger in TNFi population, and concomitant csDMARDs usage (more frequent in TNFi) (Table S3). Of note, the sample size was consequently reduced to 46 BARI, 225 TNFi and 66 OMA.

1
2
3 When analysing only these b/tsDMARD-naïve patients, both the non-adjusted (Figure 2) and
4 the adjusted differences between BARI and TNFi became larger (HR TNFi vs BARI = 2.5; 95% CI
5 [1.23 – 5.16]; p=0.01), but the differences between baricitinib and OMA group remained not
6 significantly different (HR OMA vs BARI = 1.90; 95% CI [0.71 – 5.1]; p=0.2).
7
8
9
10
11

12 **Time to discontinuation for adverse events or ineffectiveness**

13
14
15 A secondary outcome was the cumulative incidence of drug discontinuation by specific
16 reasons for discontinuation (ineffectiveness or adverse events, Figure 3). Using Fine-Gray
17 adjusted approach, we found no difference in the incidence of adverse event comparing BARI
18 to TNFi (HR = 1.46; 95% CI [0.83 - 2.57]; p=0.13), or BARI to OMA (HR = 1.34; 95% CI [0.74 -
19 2.42]; p=0.25). The incidence of drug discontinuation for ineffectiveness was more frequent
20 in TNFi compared to BARI (HR = 1.49; 95% CI [1.03 – 2.15]; p=0.01), but similar between OMA
21 and BARI (HR = 1.09; 95% CI [0.72 – 1.64]; p=0.69).
22
23
24
25
26
27
28
29
30
31
32

33 **Remission and low disease activity at 12 months**

34
35 The estimated 12-month rates of REM and LDA, estimated using CARRAC did not differ
36 significantly between the 3 groups (Figure 4). LDA ranged from 62% to 71% and REM ranged
37 from 17% to 26%.
38
39
40
41
42

43 **DISCUSSION**

44
45 In this study, the overall drug maintenance of BARI was significantly longer compared to TNFi,
46 despite the fact that it was prescribed to older patients, with longer disease duration, and
47 more previous treatment failures similar to what was observed in RA-BE-REAL, another real-
48 world study.[23] However, the adjusted 12-month response rates in terms of LDA and REM
49 did not differ significantly between BARI, TNFi and OMA groups. The difference in drug
50 discontinuation owes mainly to more treatment discontinuations for ineffectiveness in the
51
52
53
54
55
56
57
58
59
60

1
2
3 TNFi group compared to the BARI group, while drug discontinuation due to adverse event did
4
5 not differ significantly between the groups.
6
7

8 Our results are in line with previous findings comparing other JAK-inhibitors (JAKi) (i.e.
9
10 tofacitinib as well as BARI) to TNFi and OMA medications,[22,32] which reported a longer drug
11
12 maintenance of tsDMARD compared to TNFi, and similar maintenance to other bDMARDs. Of
13
14 note, Lauper et al., using data from 19 national registers, found no difference in retention time
15
16 between JAK-inhibitors and TNFi.[33] Still, Lauper et al. grouped all JAKi together in their
17
18 study, thus it is not clear if these observations remain true for BARI alone, which might differ
19
20 from other JAKi. For instance, Barbulescu et al. reported a higher drug maintenance for BARI
21
22 as compared to tofacitinib.[21]
23
24
25
26

27 It was previously shown that BARI is more efficient in relieving pain as compared to
28
29 adalimumab therapy [34] and some molecular mechanisms relevant to JAK-STAT signalling
30
31 have been hypothesized.[35] This observation has been hypothesised to result anti-
32
33 nociceptive effect independent from inflammation.[35] This faster pain relief could partially
34
35 explain why BARI has increased maintenance than other medication in our study, even though
36
37 having similar 12-months LDA and REM rates. An alternative hypothesis is that the more
38
39 convenient oral administration encourages patients to stay on medication longer. Yet, a third
40
41 possible interpretation is that patients who experienced numerous treatment failures tend to
42
43 stay on their latest therapy; however, our study accounts for this potential bias, by performing
44
45 a sensitivity analysis in a subgroup of b-tsDMARD naïve patients, which showed a similar
46
47 result. Finally, given recent discussion regarding tofacitinib safety,[36] future research needs
48
49 to clarify whether a class effect for JAKi related adverse events exist. In this analysis, we found
50
51 no indication for an increased incidence of adverse-related treatment discontinuation with
52
53
54
55
56
57
58
59
60

1
2
3 BARI compared to alternative bDMARDs. Randomized controlled trials are ongoing to further
4
5 compare safety profile of BARI versus TNFi (NCT04086745 and NCT03915964).
6
7

8 **Limitations and Strengths**

9

10 This work has several limitations, mostly inherent to the observational setting. First, as this is
11
12 a non-randomized study, we cannot formally exclude unmeasured confounding between the
13
14 groups. The available baseline variables were, in most cases, adequately balanced, except for
15
16 age. When we restricted the analysis to the subgroup of b/tsDMARD naïve patients, we found
17
18 largely similar results. Despite being limited by the small sample size, this exploratory
19
20 subgroup analysis suggests that confounding by line of treatment was adequately accounted
21
22 for in the adjusted analysis.
23
24
25

26
27 Secondly, the average length of follow-up was only approximately 200 days per TC (Figure
28
29 S2). Indeed, our study covers about 2 and a half years, and we only included TC newly initiated
30
31 during this time-windows. Also, because of the study setting, as much as 65% of TC did not
32
33 have CDAI scores recorded at the date of initiation, and many were missing at the 12-month
34
35 exact timepoint (Figure S6). Hence, our analysis of response rates relied heavily on linear
36
37 interpolation techniques, using other available timepoints, which results in large confidence
38
39 intervals for estimated response rates.[30]
40
41
42
43

44 The main strength of the study is that it relies on real-world data and includes a relatively large
45
46 number of patients providing adequate statistical power (Figure S7). As these patients are
47
48 mostly treated by office-based rheumatologists, our study population is representative of
49
50 routine clinical practice. Also, sub-group analyses and sensitivity analyses were consistent with
51
52 the main results.
53
54
55
56
57
58
59
60

CONCLUSIONS

In this non-randomized cohort study, drug maintenance of BARI was significantly higher than TNFi. However, we found no difference in drug maintenance when comparing BARI with other bDMARDs. Based on available data, the estimated 12-month response rates did not significantly differ between BARI, TNFi and OMA groups. We found no difference in treatment discontinuation for adverse event between the three groups. Overall, our results are in line with findings from randomized trials..

OTHERS

Contributorship statement

Benoît GILBERT contributed to data-management, data-analysis, figures, manuscript drafting.

Denis MONGIN contributed to data-management, data-analysis, figures, manuscript-revision.

Romain AYMON contributed to data-analysis (in particular, sensitivity analyses), manuscript revision. Kim LAUPER took part in data-analysis and manuscript revision. Cédric LAEDERMANN was involved in the study design and manuscript revision. Clémentine PERRIER was involved in the study design and manuscript revision. Ruediger MUELLER contributed to the study design and manuscript revision. Delphine COURVOISIER was involved in the study design, data-analysis and interpretation, manuscript revision. Axel FINCKH was in charge of the study design (principal investigator), data analysis, data interpretation and manuscript revision.

Competing interests

Benoît GILBERT has been once a paid speaker (Eli Lilly) and participated in advisory board (Janssen). Clementine PERRIER is employed by Eli Lilly and holds stock options (Eli Lilly and Company). Cedric LAEDERMANN is employed by Eli Lilly and holds stock options (Eli Lilly and Novartis). Axel FINCKH has received grants or contracts (Eli-Lilly, Pfizer, Abbvie, Gilead, BMS), consulting fees (Astra-Zeneca, Abbvie, Pfizer, Gilead), honorary payments (BMIS, Abbvie, Eli Lilly, Pfizer, MSD), and participated in advisory boards (Astra-Zeneca, Gilead, Novartis, Abbvie, Eli Lilly, Pfizer, J&J, Mylan, UCB). Denis MONGIN, Romain Aymon, Rüdiger Müller and Delphine COURVOISIER have no conflicts of interest to disclose.

Funding

This analysis has been made possible by financial support of Eli Lilly (Suisse) SA to the Geneva University Hospitals (HUG). Also, the SCQM is financially supported by pharmaceutical

1
2
3 industries and donors, including Eli Lilly. A list of financial supporters can be found on
4
5 www.scqm.ch/sponsors .
6
7

8 **Data sharing statement**

9

10 Restrictions apply to the availability of these data. Data is owned by a third party, the Swiss
11
12 Clinical Quality Management in Rheumatic Diseases (SCQM) foundation. Data may be
13
14 obtained after approval and permission from this license holder (SCQM). Contact information
15
16 for data request: scqm@hin.ch
17
18
19

20 **Ethical Review and Regulatory Considerations**

21

22
23 This observational study has been approved by the Geneva ethical review boards (ERBs) as
24
25 required by local law (Project ID: 2019-00930 ; approval date 28 May 2019). Every participant
26
27 has signed an information and consent form at inclusion in the SCQM registry. Hence, this
28
29 study has been conducted in accordance with the ethical principles of the Declaration of
30
31 Helsinki and is consistent with Good Pharmacoepidemiology Practices (GPPs).
32
33
34

35 **Acknowledgements**

36

37
38 Almut Scherer, from the SCQM. Also, all of the participating physicians of the SCQM in
39
40 Rheumatic Diseases. A complete list of rheumatology offices and hospitals that are
41
42 contributing to the SCQM registries can be found on www.scqm.ch/institutions .
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

- 1 Emery P, Seto Y. Role of biologics in early arthritis. *Clin Exp Rheumatol*. 2003;21:S191-194.
- 2 Lipsky PE, van der Heijde DMFM, St. Clair EW, *et al*. Infliximab and Methotrexate in the Treatment of Rheumatoid Arthritis. *New England Journal of Medicine*. 2000;343:1594–602.
- 3 Breedveld FC, Emery P, Keystone E, *et al*. Infliximab in active early rheumatoid arthritis. *Annals of the Rheumatic Diseases*. 2004;63:149–55.
- 4 Klareskog L, van der Heijde D, de Jager JP, *et al*. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *The Lancet*. 2004;363:675–81.
- 5 Weinblatt ME, Kremer JM, Bankhurst AD, *et al*. A Trial of Etanercept, a Recombinant Tumor Necrosis Factor Receptor:Fc Fusion Protein, in Patients with Rheumatoid Arthritis Receiving Methotrexate. *New England Journal of Medicine*. 1999;340:253–9.
- 6 Maini R, St Clair EW, Breedveld F, *et al*. Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. *The Lancet*. 1999;354:1932–9.
- 7 Breedveld FC, Weisman MH, Kavanaugh AF, *et al*. The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis & Rheumatism*. 2006;54:26–37.
- 8 Moreland LW, Schiff MH, Baumgartner SW, *et al*. Etanercept Therapy in Rheumatoid Arthritis. *Annals of Internal Medicine*. 1999;130:478–86.
- 9 Dixon WG, Carmona L, Finckh A, *et al*. EULAR points to consider when establishing, analysing and reporting safety data of biologics registers in rheumatology. *Annals of the Rheumatic Diseases*. 2010;69:1596–602.
- 10 Wolfe F, Michaud K. The Hawthorne Effect, Sponsored Trials, and the Overestimation of Treatment Effectiveness. *The Journal of Rheumatology*. 2010;37:2216–20.
- 11 Wolfe F, Michaud K. The loss of health status in rheumatoid arthritis and the effect of biologic therapy: a longitudinal observational study. *Arthritis Res Ther*. 2010;12:R35.
- 12 Østergaard M, Unkerskov J, Linde L, *et al*. Low remission rates but long drug survival in rheumatoid arthritis patients treated with infliximab or etanercept: results from the nationwide Danish DANBIO database. *Scandinavian Journal of Rheumatology*. 2007;36:151–4.

- 13 Souto A, Maneiro JR, Gómez-Reino JJ. Rate of discontinuation and drug survival of biologic therapies in rheumatoid arthritis: a systematic review and meta-analysis of drug registries and health care databases. *Rheumatology*. 2016;55:523–34.
- 14 Taylor PC, Keystone EC, van der Heijde D, *et al*. Baricitinib versus Placebo or Adalimumab in Rheumatoid Arthritis. *New England Journal of Medicine*. 2017;376:652–62.
- 15 Fleischmann R, Schiff M, Heijde D van der, *et al*. Baricitinib, Methotrexate, or Combination in Patients With Rheumatoid Arthritis and No or Limited Prior Disease-Modifying Antirheumatic Drug Treatment. *Arthritis & Rheumatology*. 2017;69:506–17.
- 16 Dougados M, van der Heijde D, Chen Y-C, *et al*. Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. *Ann Rheum Dis*. 2017;76:88–95.
- 17 Genovese MC, Kremer J, Zamani O, *et al*. Baricitinib in Patients with Refractory Rheumatoid Arthritis. *N Engl J Med*. 2016;374:1243–52.
- 18 Smolen JS, Genovese MC, Takeuchi T, *et al*. Safety Profile of Baricitinib in Patients with Active Rheumatoid Arthritis with over 2 Years Median Time in Treatment. *J Rheumatol*. 2019;46:7–18.
- 19 Keystone EC, Taylor PC, Tanaka Y, *et al*. Patient-reported outcomes from a phase 3 study of baricitinib versus placebo or adalimumab in rheumatoid arthritis: secondary analyses from the RA-BEAM study. *Ann Rheum Dis*. 2017;76:1853–61.
- 20 Taylor PC, Takeuchi T, Burmester GR, *et al*. Safety of baricitinib for the treatment of rheumatoid arthritis over a median of 4.6 and up to 9.3 years of treatment: final results from long-term extension study and integrated database. *Ann Rheum Dis*. 2022;81:335–43.
- 21 Barbulescu A, Askling J, Chatzidionysiou K, *et al*. Effectiveness of baricitinib and tofacitinib compared with bDMARDs in RA: results from a cohort study using nationwide Swedish register data. *Rheumatology*. 2022;keac068.
- 22 Pappas DA, John GS, Etzel CJ, *et al*. Comparative effectiveness of first-line tumour necrosis factor inhibitor versus non-tumour necrosis factor inhibitor biologics and targeted synthetic agents in patients with rheumatoid arthritis: results from a large US registry study. *Annals of the Rheumatic Diseases*. 2021;80:96–102.
- 23 Alten R, Burmester GR, Cerinic MM, *et al*. Pos0666 a Multinational, Prospective, Observational Study in Patients with Rheumatoid Arthritis Receiving Baricitinib, Targeted Synthetic or Biologic Disease-Modifying Therapies: 12 Month Time to Discontinuation, Effectiveness and Patient Reported Outcome Data from the European Cohort. *Annals of the Rheumatic Diseases*. 2022;81:606–7.
- 24 Pan SMD, Scherer A, Gabay C, *et al*. Differential drug retention between anti-TNF agents and alternative biological agents after inadequate response to an anti-TNF agent in rheumatoid arthritis patients. *Annals of the Rheumatic Diseases*. 2012;71:997–9.

- 1
2
3 25 Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease
4 Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis. ;9.
5
6
7 26 Lauper K, Finckh A. Predictive factors of treatment persistence in rheumatoid arthritis.
8 *Joint Bone Spine*. 2020;87:531–4.
9
10 27 White H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test
11 for Heteroskedasticity. *Econometrica*. 1980;48:817–38.
12
13 28 Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing
14 Risk. *Journal of the American Statistical Association*. 1999;94:496–509.
15
16
17 29 Mongin D, Lauper K, Finckh A, *et al*. Accounting for missing data caused by drug cessation
18 in observational comparative effectiveness research: a simulation study. *Ann Rheum Dis*.
19 2022;annrheumdis-2021-221477.
20
21 30 Mongin D, Lauper K, Turesson C, *et al*. Imputing missing data of function and disease
22 activity in rheumatoid arthritis registers: what is the best technique? *RMD Open*.
23 2019;5:e000994.
24
25
26 31 Buuren S van, Groothuis-Oudshoorn K. Mice : Multivariate Imputation by Chained
27 Equations in R. *J Stat Soft*. 2011;45. doi: 10.18637/jss.v045.i03
28
29
30 32 Finckh A, Tellenbach C, Herzog L, *et al*. Comparative effectiveness of antitumour necrosis
31 factor agents, biologics with an alternative mode of action and tofacitinib in an
32 observational cohort of patients with rheumatoid arthritis in Switzerland. *RMD Open*.
33 2020;6:e001174.
34
35
36 33 Lauper K, Iudici M, Mongin D, *et al*. Effectiveness of TNF-inhibitors, abatacept, IL6-
37 inhibitors and JAK-inhibitors in 31 846 patients with rheumatoid arthritis in 19 registers
38 from the 'JAK-pot' collaboration. *Annals of the Rheumatic Diseases*. 2022;81:1358–66.
39
40
41 34 Taylor PC, Lee YC, Fleischmann R, *et al*. Achieving Pain Control in Rheumatoid Arthritis
42 with Baricitinib or Adalimumab Plus Methotrexate: Results from the RA-BEAM Trial. *JCM*.
43 2019;8:831.
44
45
46 35 Busch-Dienstfertig M, González-Rodríguez S. IL-4, JAK-STAT signaling, and pain. *JAKSTAT*.
47 2013;2:e27638.
48
49
50 36 Álvaro-Gracia JM, García-Llorente JF, Valderrama M, *et al*. Update on the Safety Profile of
51 Tofacitinib in Rheumatoid Arthritis from Clinical Trials to Real-World Studies: A Narrative
52 Review. *Rheumatol Ther*. 2021;8:17–40.
53
54
55
56
57
58
59
60

FIGURE LEGENDS

Figure 1 - Non-adjusted time to drug discontinuation analyses (Kaplan-Meier), SCQM, 2017-2020.

These “survival curves” represent the drug maintenance after initiation, as the estimated proportion of patients still on therapy, by treatment group. Death and loss to follow-up were censored. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs. Log-Rank BARI vs TNFi : $p < 0.001$. Log-Rank BARI vs OMA: $p = 0.11$.

Figure 2 - Unadjusted time to drug discontinuation in b/tsDMARD-naïve patients, SCQM, 2017-2020.

These Kaplan-Meier curves represent the crude “survival” of drug prescription, by treatment group. Death and loss to follow-up are censored. BARI = baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs. Log-Rank BARI vs TNFi : $p = 0.003$. Log-Rank BARI vs OMA : $p = 0.15$.

Figure 3 - Cumulative incidence of drug discontinuation by stop reason and by type of treatment, SCQM, 2017-2020.

This figure represents the unadjusted cumulative incidence of drug discontinuation, by group and by reason of discontinuation. BARI = baricitinib. TNFi = Tumor Necrosis Factor. OMA = Other mode of Action bDMARDs.

Figure 4 - Estimated response rates at 12-months (CARRAC), SCQM, 2017-2020.

BARI = baricitinib. TNFi = Tumor Necrosis Factors Inhibitors. OMA = Other Mode of Action bDMARDs. LDA = Low Disease Activity (i.e. CDAI score ≤ 10), in light grey. REM = Remission

1
2
3 (i.e. CDAI score ≤ 2.8), in dark grey. 95% CI are represented. This method does not allow
4
5 computing p-values. Nb: two estimates were obtained in the BARI group and averaged to
6
7 display only one representative value on the plot. Actual row output was 68% (95%CI [55% ;
8
9 80%]) (BARI vs TNFi model) or 62% (95% CI [54% ; 70%]) (BARI vs OMA model) for LDA, and
10
11 23% (95% CI [14% ; 31%]) (BARI vs TNFi model) or 17 % (95% CI [10% ; 24%]) (BARI vs OMA
12
13 model) for REM.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

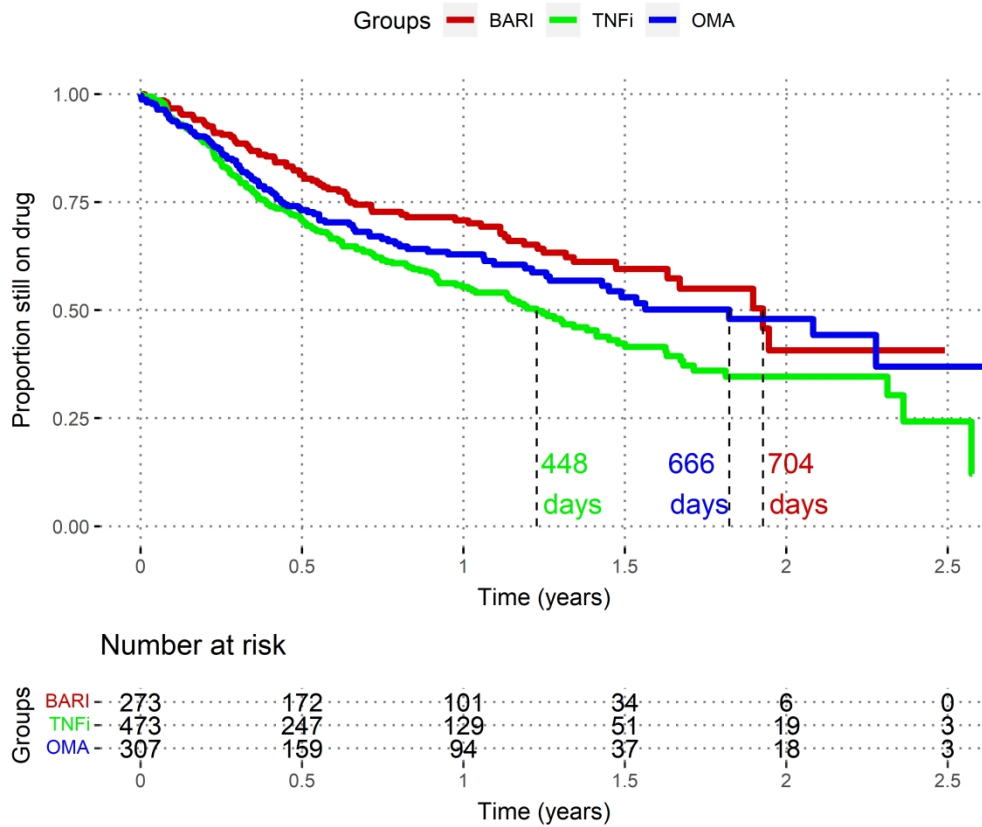


Figure 1 - Non-adjusted time to drug discontinuation analyses (Kaplan-Meier), SCQM, 2017-2020. These "survival curves" represent the drug maintenance after initiation, as the estimated proportion of patients still on therapy, by treatment group. Death and loss to follow-up were censored. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs. Log-Rank BARI vs TNFi : $p < 0.001$. Log-Rank BARI vs OMA: $p = 0.11$.

451x386mm (197 x 197 DPI)

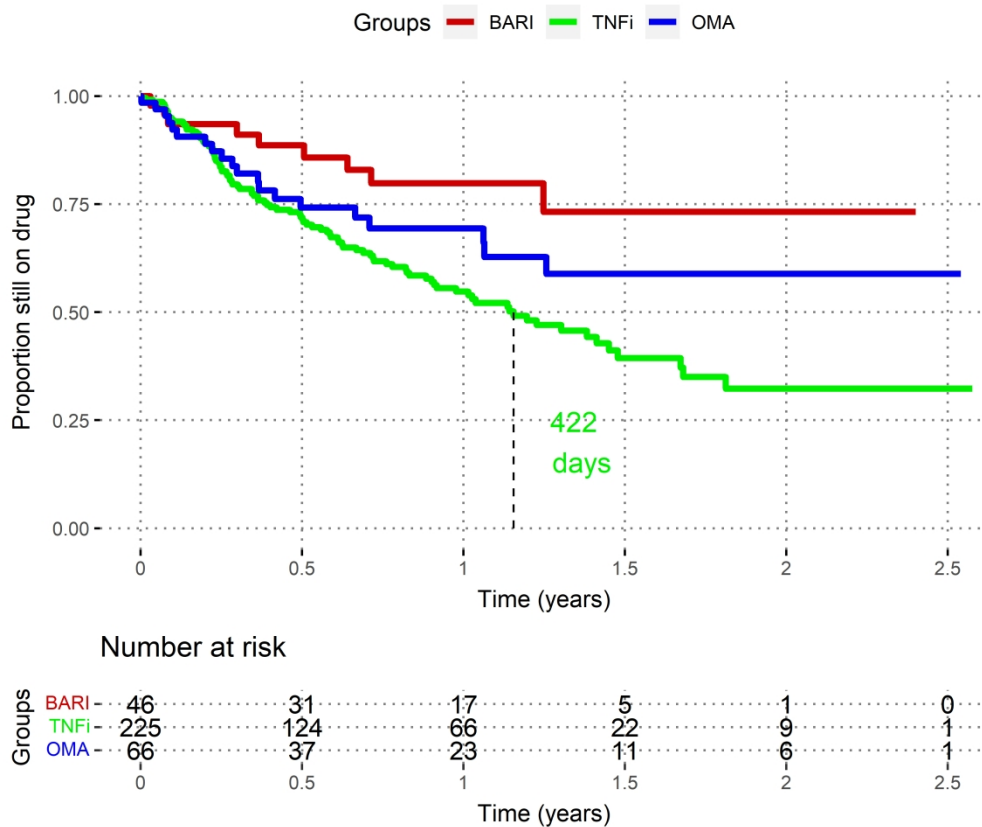


Figure 2 - Unadjusted time to drug discontinuation in b/tsDMARD-naïve patients, SCQM, 2017-2020. These Kaplan-Meier curves represent the crude “survival” of drug prescription, by treatment group. Death and loss to follow-up are censored. BARI = baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Mode of Action bDMARDs. Log-Rank BARI vs TNFi : $p = 0.003$. Log-Rank BARI vs OMA : $p = 0.15$.

451x386mm (197 x 197 DPI)

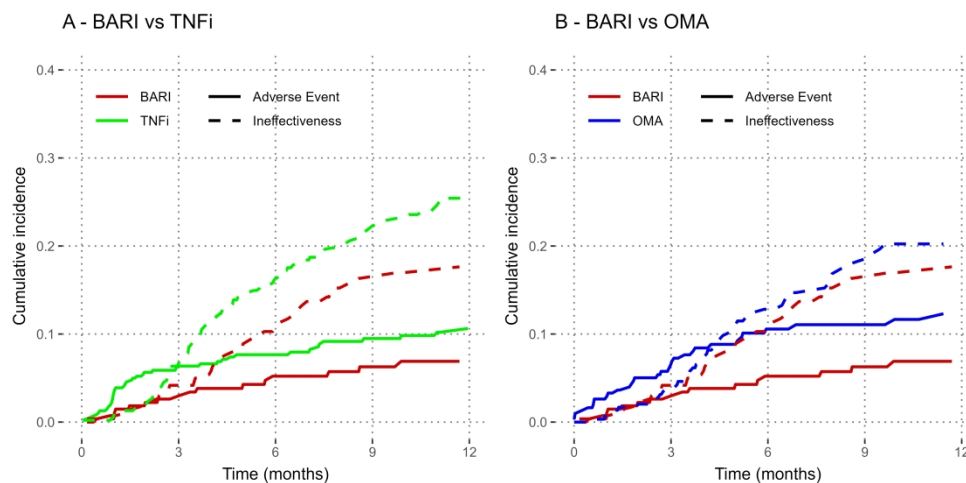


Figure 3 - Cumulative incidence of drug discontinuation by stop reason and by type of treatment, SCQM, 2017-2020. This figure represents the unadjusted cumulative incidence of drug discontinuation, by group and by reason of discontinuation. BARI = baricitinib. TNFi = Tumor Necrosis Factor. OMA = Other mode of Action bDMARDs.

644x322mm (197 x 197 DPI)

REM and LDA rates
by type of treatment
(CARRAC)

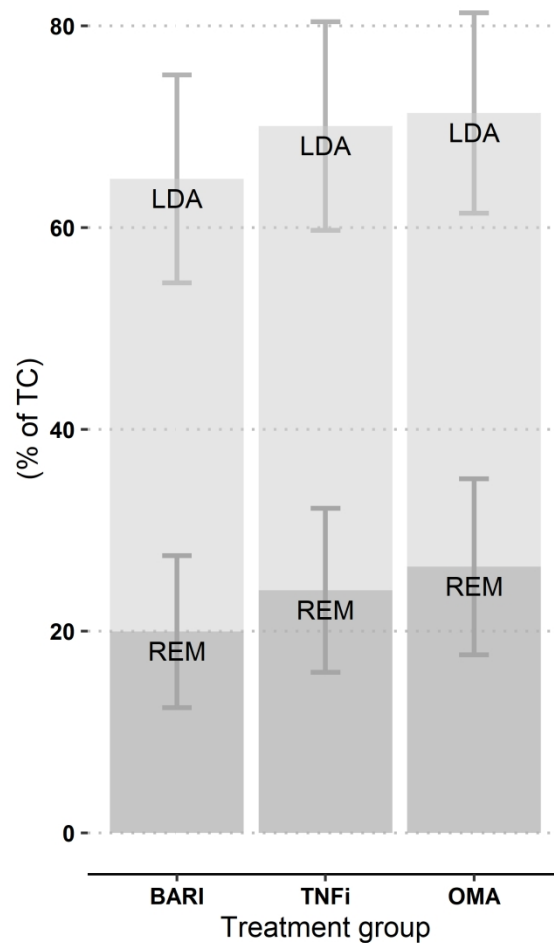


Figure 4 - Estimated response rates at 12-months (CARRAC), SCQM, 2017-2020. BARI = baricitinib. TNFi = Tumor Necrosis Factors Inhibitors. OMA = Other Mode of Action bDMARDs. LDA = Low Disease Activity (i.e. CDAI score ≤ 10), in light grey. REM = Remission (i.e. CDAI score ≤ 2.8), in dark grey. 95% CI are represented. This method does not allow computing p-values. Nb: two estimates were obtained in the BARI group and averaged to display only one representative value on the plot. Actual row output was 68% (95%CI [55% ; 80%]) (BARI vs TNFi model) or 62% (95% CI [54% ; 70%]) (BARI vs OMA model) for LDA, and 23% (95% CI [14% ; 31%]) (BARI vs TNFi model) or 17% (95% CI [10% ; 24%]) (BARI vs OMA model) for REM.

257x451mm (197 x 197 DPI)

SUPPLEMENTARY DATA

Selection of eligible treatment courses

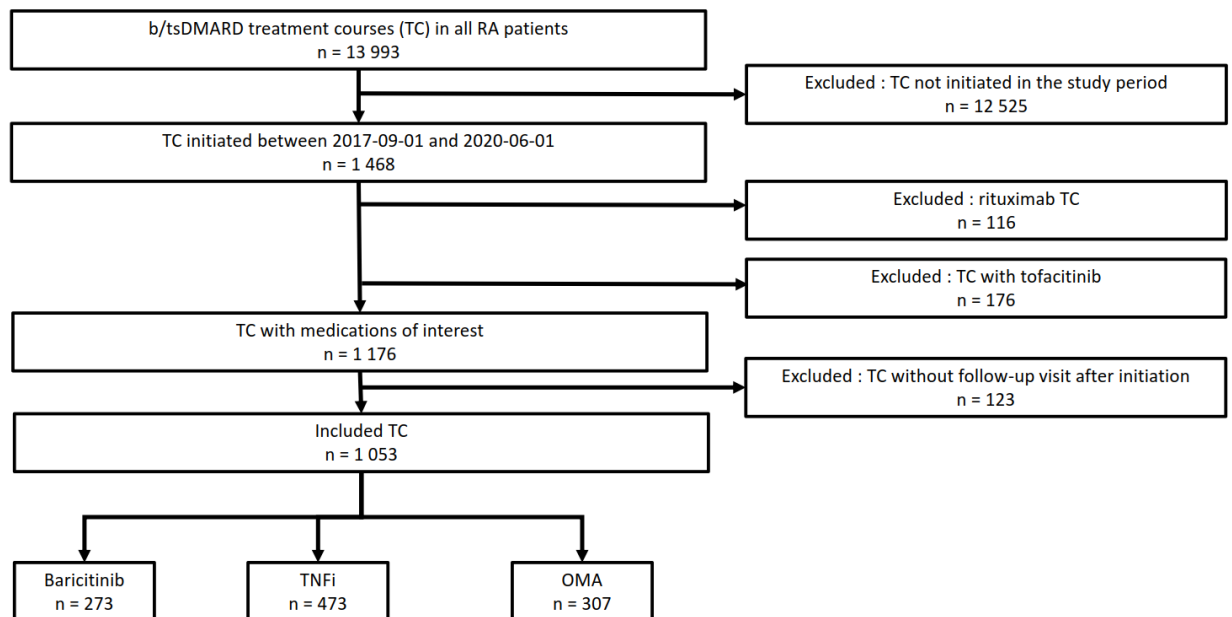


Figure S1 - Selection of eligible treatment courses, SCQM, 2017-2020.

Selection of Treatment courses included in final analysis. TC = Treatment Courses. RA = Rheumatoid Arthritis. bDMARD = biological DMARD. tsDMARD = targeted synthetic DMARD. b/tsDMARD = biological and/or targeted synthetic DMARD. TNFi = TNF inhibitors. OMA = Other Mode of Action bDMARDs.

Notice on TC duration

Due to frequent changes in medication and short study period, it has to be underlined that the median duration of a TC approximates 200 days. The proportion of TC with follow-up data of at least one year is 37% for BARI, 27% for TNFi and 31% for OMA (Figure S2) - i.e. most TC were started less than 12 months before the date of data extraction.

Notice this % is different from the % of patient still under therapy that we estimate using Kaplan-Meier or Cox model. Indeed, the latter includes a censoring of the lost-to-follow-up patients, hence the denominator is different. As a consequence, this does not contradict the reported “median prescription survival timey”, for instance of 704 days for BARI TCs. The latter is the output estimated by the Kaplan-Meier model, taking censoring into account; it does not imply that actual observations in the dataset have this duration.

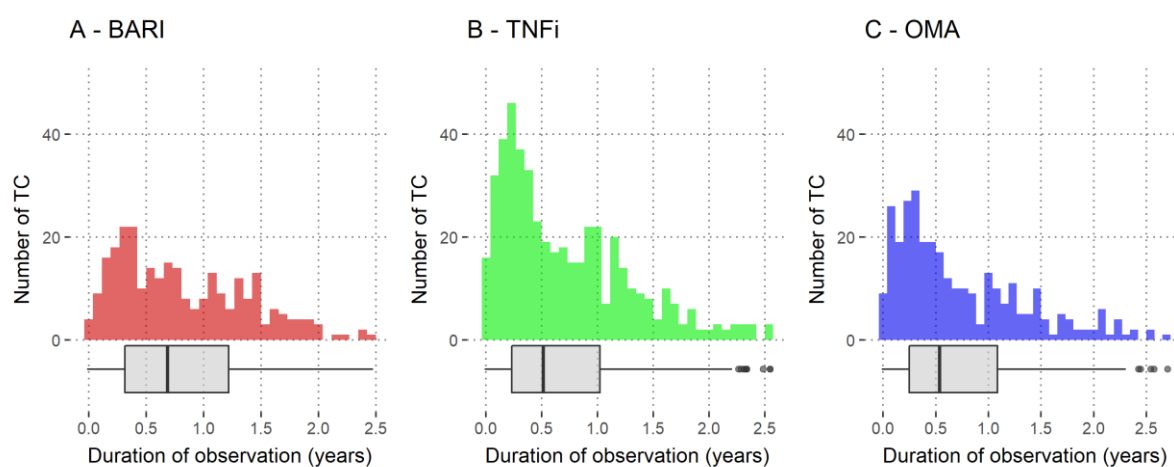


Figure S2: Distribution of the observation time for included TCs, per group, SCQM, 2017-2020.

Most of the treatment courses have an actual duration and/or follow-up period of less than one year. TC = Treatment Course. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors.

OMA = Other Mode of Action bDMARDs.

Variable definitions

Below we give additional detail about included covariates:

Age: age in years, at TC initiation. Continuous variable.

Gender: male or female. Categorical variable.

BMI: BMI at TC initiation. Continuous variable.

CDAI score: CDAI score at TC initiation. Continuous variable. If missing, imputed according to procedure described in methods section.

Disease duration: time interval between RA diagnosis date and TC initiation date. Continuous variable, expressed in years, but used in decades in models.

Smoking status: smoking status at TC initiation. Categorical variable (current-, former-, never-smoker).

Concomitant csDMARD: yes/no variable. A concomitant csDMARD was defined as csDMARD prescription ongoing for at least 40% of the duration of the TC. Otherwise, the TC was categorized as monotherapy. csDMARDs included: methotrexate, sulfasalazin, leflunomide, azathioprine and hydroxy-chloroquine, alone or in combination.

Concomitant glucocorticoid: Yes/no variable. Concomitant glucocorticoid usage was defined as having at least one active prescription of glucocorticoid, at any dose, at any timepoint of the TC.

Line of therapy: strictly speaking, this categorical variable is displaying: [*number of previous TC ever* + 1]. 4 or more has been grouped in the same category. Hence, it is considering all data of the SCQM registry, i.e. TCs initiated before our study period are also accounted for as previous therapies.

Seropositivity: yes/no variable. Seropositivity is defined as positivity for anti-citrullinated peptide antibodies and/or rheumatoid factor.

Time to all cause discontinuation

Table S1: Crude treatment discontinuation by group and by reason, SCQM, 2017-2020.

	BARI (TCs = 273; 273 pts)	TNFi (TCs = 473; 408 pts)	OMA (TCs = 307; 289 pts)
Treatment discontinuation (all causes)	30 %	43 %	35 %
For adverse events	8 %	10 %	10 %
For ineffectiveness	16 %	23 %	17 %
For remission	0 %	1 %	0 %
For other reason	5 %	8 %	7 %
<p>Table S1 legend: % are computed on total number of TCs per group, for the whole study period. BARI = baricitinib. TNFi = TNF inhibitors. OMA = Other Mode of Action drugs. TC = Treatment Courses. Pts = Number of patients. Due to rounding, the sum of the percentages of the causes of discontinuation may not correspond exactly to the total treatment discontinuation percentage.</p>			

Cox model output

Table S2 contains the complete output of the two adjusted Cox models used in the main time-to-drug discontinuation analysis.

	BARI vs TNFi			BARI vs OMA		
	Hazard ratio	95% CI	p	Hazard ratio	95% CI	p
TNFi (vs baricitinib)	1.76	1.32-2.35	<0.001	-	-	-
OMA (vs baricitinib)	-	-	-	1.27	0.93-1.72	0.13
Adjusting variables:						
Age (decades)	1.03	0.92-1.14	0.61	0.98	0.86-1.10	0.69
BMI	1.01	0.98-1.04	0.51	0.98	0.94-1.02	0.31
TC with csDMARD	0.84	0.66-1.09	0.19	1.22	0.90-1.67	0.20
Glucocorticoid usage	1.29	0.93-1.79	0.12	1.86	1.32-2.61	<0.001
CDAI score	1.40	1.26-1.56	<0.001	1.15	1.03-1.28	0.01
Disease duration (decades)	0.95	0.81-1.10	0.46	0.85	0.70-1.03	0.10
Current smoker (vs non-smoker)	1.20	0.86-1.68	0.28	1.09	0.73-1.64	0.66
Ever smoker (vs non-smoker)	1.10	0.79-1.52	0.57	1.38	0.95-2.00	0.09
2nd line therapy (vs 1st)	1.11	0.80-1.53	0.52	1.37	0.81-2.33	0.24
3rd line therapy (vs 1st)	0.98	0.64-1.51	0.93	1.56	0.92-2.64	0.10
4th or later line (vs 1st)	1.06	0.75-1.51	0.73	1.57	0.93-2.63	0.09
Female gender	1.05	0.78-1.42	0.74	1.16	0.81-1.67	0.41
Seropositivity	0.77	0.59-1.01	0.055	0.94	0.67-1.31	0.71

BARI = baricitinib. TNFi = TNF inhibitors. OMA = Other Mode of Action drugs. CI = Confidence Interval. BMI = Body Mass Index. CDAI = Clinical Disease Activity Index.

Figure S3 below gives the exact same information as Table S2:

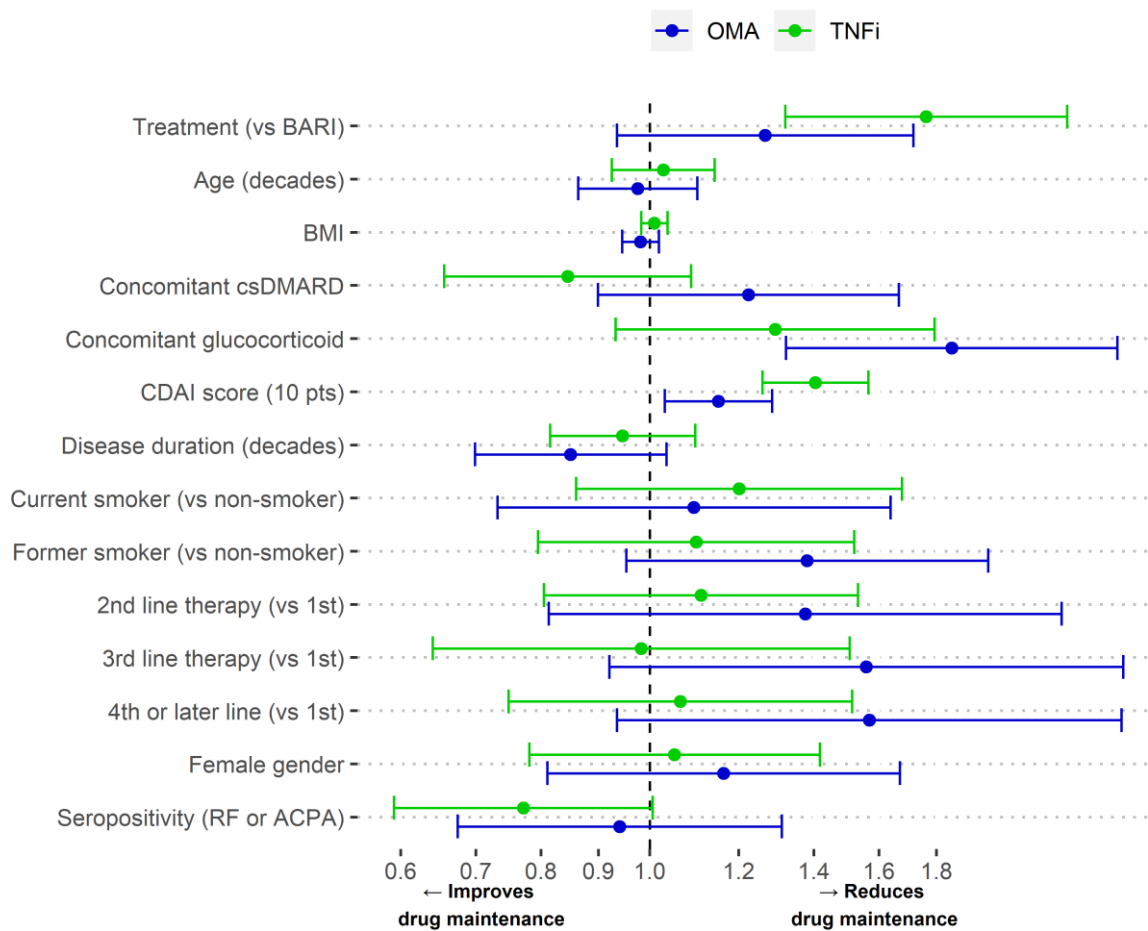


Figure S3: Hazard ratio of drug discontinuation (95% CI).

BARI = Baricitinib. TNFi = Tumor Necrosis Factors Inhibitors. OMA = Other mode of Action bDMARDs. BMI = Body Mass Index. CDAI = Clinical Disease Activity Index. RF = Rheumatoid Factor. ACPA = Anti-citrullinated Peptides Antibodies.

The corresponding cox-adjusted drug-survival curves are provided below:

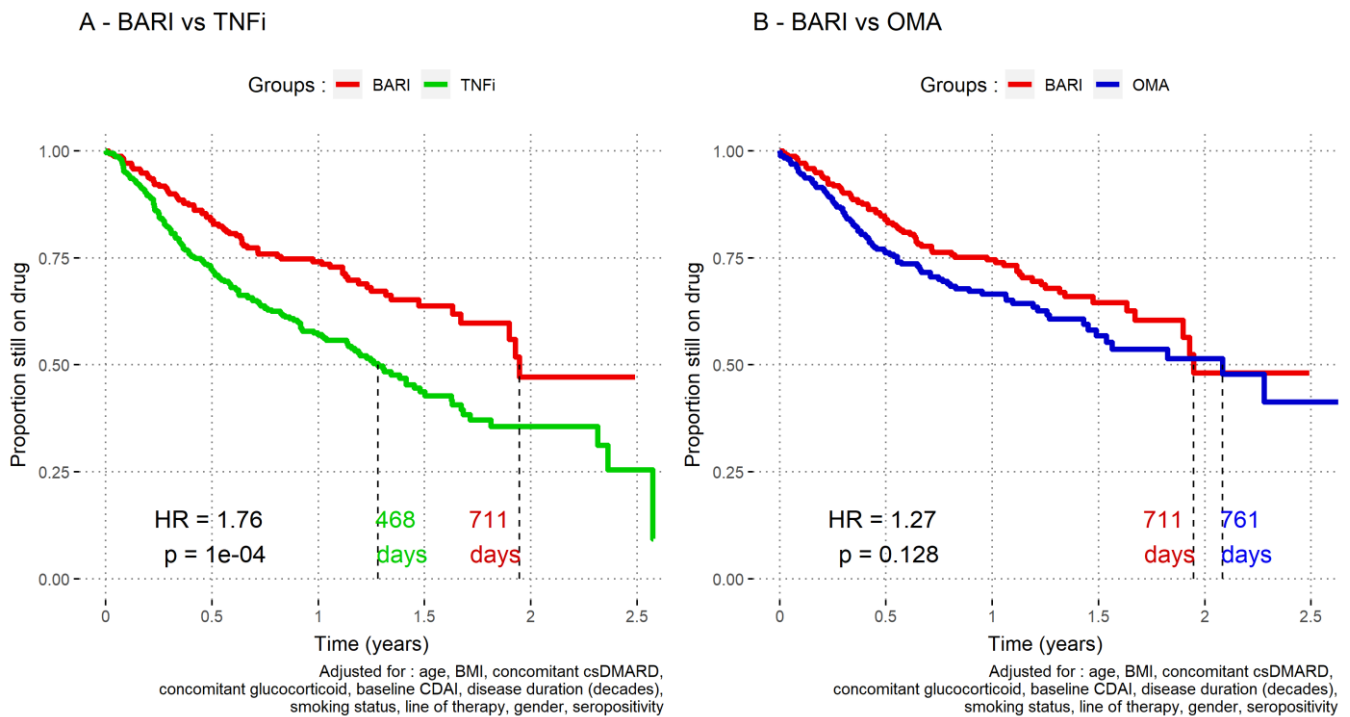


Figure S4: Multivariable Cox model of drug discontinuation by type of treatment, SCQM, 2017-2020.

These curves are merely the visualisation of Cox models presented in Table S1 and Figure S2.

BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action bDMARDs

Models are adjusted for : age, BMI, concomitant csDMARD, concomitant glucocorticoid, baseline CDAI, disease duration, smoking status, line of therapy, gender, seropositivity.

Sensitivity analysis using AIPTW

As a sensitivity analysis, the main time to drug discontinuation was also performed using “augmented inverse probability of treatment weighting” (AIPTW), including the same covariates. In other words, we combined a propensity score using a logistic regression model and an inverse probability weighted Cox regression. We used the *RiskRegression* package in R, to obtain risk ratios.

Figure S5 represents the absolute risk of treatment discontinuation, for all included timepoints. At one year, the adjusted discontinuation risk in BARI was 19 % lower than in TNFi group ($p < 0.001$) (Figure S5 A), with a risk ratio of 1.76 (95% CI [1.19 – 2.34]; $p = 0.009$). Similarly, at one year, the adjusted treatment discontinuation risk in BARI was 8 % lower than in the OMA group ($p = 0.06$) (Figure S5 B), with a risk ratio of 1.28 (95% [0.91 – 1.65]; $p = 0.14$).

Overall, this sensitivity analysis confirms the findings reported in the main body of the article.

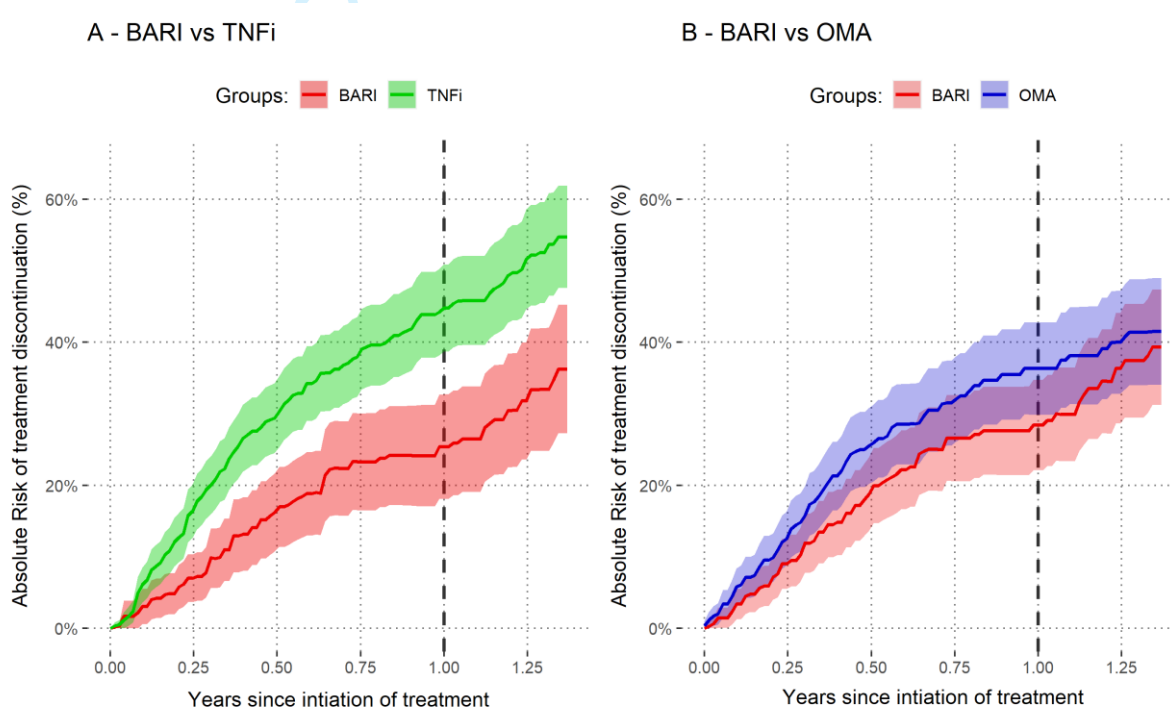


Figure S5: Absolute risk of treatment discontinuation by type of treatment (AIPTW), SCQM, 2017-2020

BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action bDMARDs.

AIPTW = Augmented Inverse Probability of Treatment Weighting. Adjusted for : age, bmi, concomitant csDMARDs, prednisone usage, baseline CDAI, disease duration, smoking status, line of therapy, gender, seropositivity.

Time to all-cause-discontinuation in b/tsDMARD-naïve patients

Table S3: Baseline characteristics of study population, b/tsDMARD-naïve patients, SCQM-RA registry, 2017-2020.

Patient-Variables	BARI (n = 46)		TNFi (n = 225)		OMA (n = 66)		p values
		Miss.		Miss.		Miss.	
Female	70 %	0	71 %	1	73 %	1	0.88
Age (years)	57 (15)	0	52 (14)	1	57 (16)	1	<0.01
Disease duration (years)	6 (6)	1	5 (7)	13	7 (9)	2	0.24
CDAI baseline (raw data)	16 (8)	31	18 (10)	135	18 (14)	42	0.77
CDAI baseline (imputed)	12 (7)	0	14 (9)	0	14 (10)	0	0.61
Obesity (BMI > 30)	11 %	13	13 %	58	5 %	27	0.28
Smoking							
Current	28 %	9	18 %	42	14 %	13	0.18
Former	26 %		24 %		21 %		
Never	26 %		39 %		46 %		
Seropositive (ACPA or RF)	80 %	1	69 %	5	76 %	2	0.20
TC variables							
Dose of BARI (4mg)	83 %	0	-	-	-	-	-
TC duration > 12-months	37 %	0	29 %	0	35 %	0	0.48
Concomitant csDMARD	41 %	0	66 %	0	50 %	0	<0.01
Line of Therapy							
1 st (= bio-naïve)							
2 nd	100 %	0	100 %	0	100 %	0	-
3 rd							
4 th or later							
Previous tsDMARD use (non-BARI)	0 %	0	0 %	0	0 %	0	-
Concomitant glucocorticoid (at any time)	13 %	0	20 %	0	17 %	0	0.50
Mean dose of concomitant glucocorticoid (mg)	1 (4)	0	3 (6)	0	2 (6)	0	0.50

Table S3: BARI = baricitinib. TC = Treatment Courses. CDAI = Clinical Disease Activity Index.

TNFi = Tumor Necrosis Factor Inhibitors. OMA = bDMARDs with Other Mode of Action.

tsDMARD = targeted synthetic DMARDs. ACPA = Anti Citrullinated Peptide Antibodies. RF

= Rheumatoid Factor. Miss. = number of missing values. p-values are computed with

either Chi² or ANOVA.

Response rates – raw CDAI data

Figure S6 shows the crude available values for CDAI scores, by type of treatment and time.

Only a minority of CDAI scores were assessed at 0- or 12-month timepoints of TCs (i.e. 680/1053 = 65% were missing for baseline value, and 908/1053 = 86% were missing for exact 12-month value). Future research would certainly benefit having CDAI scores assessed at regular and homogenous time-intervals, based on the initiation date of biological therapies.

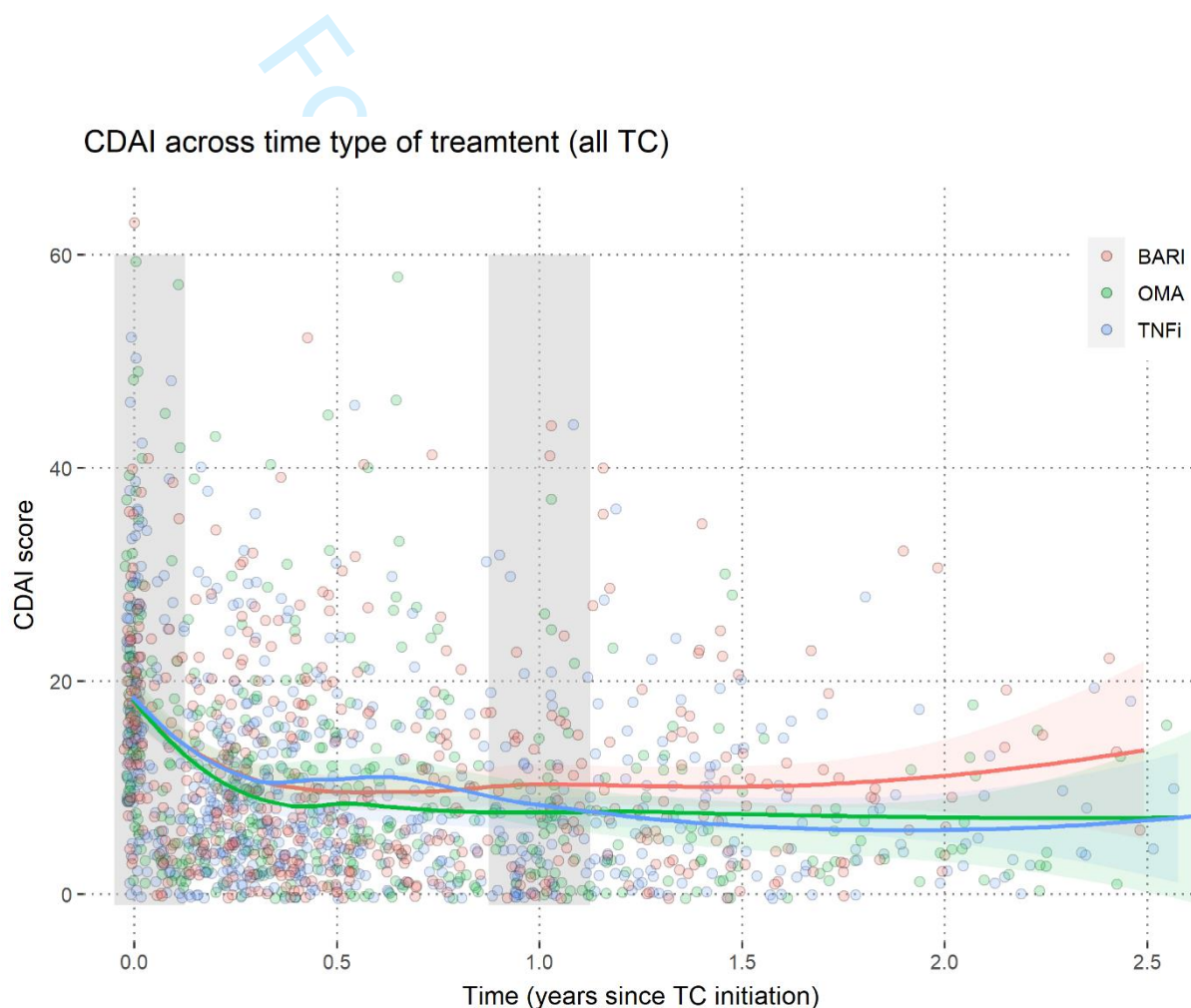


Figure S6: CDAI across time by type of treatment, raw data, SCQM, 2017-2020.

Only a minority of CDAI score were obtained sharp at 0 or 12 months of TCs. CDAI = Clinical Disease Activity Index. TC = Treatment course. BARI = Baricitinib. TNFi = Tumor Necrosis Factor Inhibitors. OMA = Other Modes of Action b/tsDMARDs.

Study size

Based on estimates from similar analyses with tofacitinib (TOFA) performed in this registry, we calculated the number of patients that would be needed to detect a significant decrease in time to all cause-discontinuation of treatment (hazard ratio) between treatment groups using the method described by Schoenfeld and Richter. We assumed a statistical power of 80%, a type I error probability of 0.05, a median BARI retention of 30 months, the inclusion of 3 patients on TNFi for every patient on BARI, an accrual time of 2 years, and additional follow-up of 6 months. We display below the sample size for the BARI group for a range of possible effect sizes (“hazard ratio” between 1.1 and 1.8).

If the true hazard ratio is similar to the one found with TOFA compared to TNFi after a single TNFi failure (HR 1.68), we will need to study 149 patients on BARI and 447 patients on TNFi to be able to reject the null hypothesis that the experimental and control curves are equal with probability (power) of 80%. Pragmatically, we propose to start the analysis of the data only once at least 200 patients on BARI have been included and followed for an average of at least 18 months.

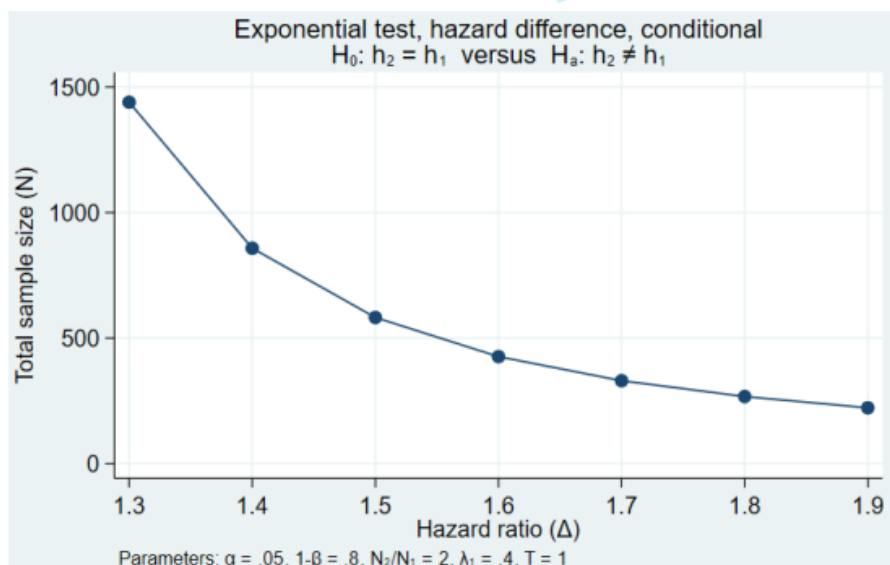


Figure S7: Estimated total sample size for two-sample comparison of survivor functions

1 - SURVIVAL ANALYSIS

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)
library(gridExtra)
library(grid)
library(cmprsk)
library(mstate)
library(cobalt)

rm(list = ls()) # To select all loaded objects and delete them
setwd(dirname(rstudioapi::getActiveDocumentContext())$path) # setting
up working directory in the location of the .Rmd file

load("./1_datamanaged_files/datamanaged.Rdata") # loading data.managed
data

Loading fonctions

# home-made function to force writing with two decimals
formattable = function(nbr){return(formatC(nbr,format = "f",digits =
nombreapresvirgule))}
nombreapresvirgule <- 2

# Home-made Fonction to write the p value (by denis)
writepvalue = function(pvalue) {
  if (is.na(pvalue)) {result <- NA} else {
    if(pvalue < 0.001) {
```



```

1
2
3     result <- "<0.001"
4   } else if (pvalue <0.01) {
5     result <- formatC( pvalue ,format = "f",digits = 3)
6   }
7   else {
8     (result <- formatC( pvalue ,format = "f",digits = 2) )
9     i = 1
10    while(result == 0.05) {
11      result <- formatC( pvalue ,format = "f",digits = 2 + i)
12      i = i + 1
13    }
14  }
15  return(result)
16 }
17 }
18 options(scipen = 999)
19
20
21

```

Mini Exploration

```

22 uniqueN(BARI_DATA[]$patient_id) # number of patients (< than number of
23 TC)
24 uniqueN(BARI_DATA[]$TC_id) # unumber of TC
25
26

```

```

27 plot <- qplot(x = BARI_DATA[]$time_on_drug)+
28   geom_vline(xintercept = 365, size = 1.2, alpha = 0.5)+
29   geom_text(aes(x = 365 + 40, label="1 Year", y=20), colour="white",
30   angle=0)+
31
32   geom_vline(xintercept = 2*365, size = 1.2, alpha = 0.5)+
33   geom_text(aes(x = 2*365 + 40, label="2 Year", y=20), colour="black",
34   angle=0)+
35
36   geom_vline(xintercept = mean(BARI_DATA$time_on_drug), color = "red",
37   size = 1.2)+
38   geom_text(aes(x = mean(BARI_DATA$time_on_drug) + 40, label="Mean",
39   y=20), colour="red", angle=0)+
40
41   geom_vline(xintercept = median(BARI_DATA$time_on_drug), color =
42   "green", size = 1.2)+
43   geom_text(aes(x = median(BARI_DATA$time_on_drug) - 40,
44   label="Median", y=20), colour="green", angle=0)+
45
46   labs(x = "Duration of TC (days)", y = "Number of TC", title =
47   "Repartition of the duration of included TC (all groups)")+
48
49   theme_pubclean()
50
51
52
53 plot
54
55 mean(BARI_DATA$time_on_drug)
56
57
58
59
60

```

```

1
2
3 median(BARI_DATA$time_on_drug)
4
5 # Nb : Research protocol said we wanted a follow-up duration of
6 "average of 18 months"
7 mean(BARI_DATA[cohort == "BARI"]$time_on_drug)/30
8 mean(BARI_DATA[cohort == "TNFi"]$time_on_drug)/30
9 mean(BARI_DATA[cohort == "OMA"]$time_on_drug)/30 # looks more like 9
10 months..
11
12
13 # ok, 24-month follow-up will be complicated
14 uniqueN(BARI_DATA[time_on_drug > 2*365, TC_id]) # number if TC with
15 duration > 24 months
16
17
18

```

1. [0] Table 1 BARI vs TNFi et OMA bDMARDs

Common table with all the data

Showing NA to have complete counts and accurate % in each category

```

24 BARI_DATA[,time_on_drugDiff0 := as.numeric(time_on_drug > 0)] # time
25 in drug < 0

```

```

26 BARI_DATA[,time_on_drug365 := as.numeric(time_on_drug > 365.25)]
27

```

```

28
29 myVars2 <- c("gender", "age_base", "disease_duration_base_years",
30 "CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
31 "seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
32 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
33 "PREDNISON_STEROID_dose", "dose", "initiation_year",
34 "time_on_drug", "HAQ_score_base")
35

```

```

36 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
37 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
38 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
39 "seropositivity_base")
40

```

```

41 nonnormalVars2 <- c()

```

```

42 tab2 <- CreateTableOne(vars = myVars2, data = BARI_DATA, factorVars =
43 catVars2, strata = "cohort", test = F, includeNA = T)

```

```

44 tablexp2 <- print(tab2, nonnormal= nonnormalVars2, catDigits = 1,
45 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
46

```

saving table 1 NA

```

49 write.xlsx(tablexp2, file =
50 "./3_clean_output/BARI_3_groups_table1_NA.xlsx")
51

```

Without NA to obtain adequate p values

```

54 BARI_DATA[,time_on_drugDiff0 := as.numeric(time_on_drug > 0)] # time
55 in drug < 0
56
57
58
59

```

```

1
2
3 BARI_DATA[,time_on_drug365 := as.numeric(time_on_drug > 365.25)]
4
5 myVars2 <- c("gender", "age_base", "disease_duration_base_years",
6 "CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
7 "seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
8 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
9 "PREDNISON_STEROID_dose", "dose", "initiation_year",
10 "time_on_drug", "HAQ_score_base")
11
12
13 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
14 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
15 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
16 "seropositivity_base")
17
18 nonnormalVars2 <- c()
19 tab2 <- CreateTableOne(vars = myVars2, data = BARI_DATA, factorVars =
20 catVars2, strata = "cohort", test = T, includeNA = F)
21 tablexp2 <- print(tab2, nonnormal= nonnormalVars2, catDigits = 1,
22 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
23
24

```

Saving table 1

```

25
26 write.xlsx(tablexp2, file =
27 "./3_clean_output/BARI_3_groups_table1.xlsx")
28

```

But BMJ-Open reviewer 2 aksed for p-values in Table 1 that account for patients providing multiple TCs. Here is how to proceed (it is a bit less conservative):

```

29
30
31
32 library(lme4)
33 library(lmerTest)
34

```

```

35
36 # Two glmer() models have to be compared, to assess the impact of
37 grouping, for each baseline variable).
38

```

```

39 # gender
40 gender.tab1 <- glmer(gender ~ cohort + (1|patient_id), data =
41 BARI_DATA , family = "binomial")
42 gender.null <- glmer(gender ~ (1|patient_id), data = BARI_DATA ,
43 family = "binomial")
44 anova(gender.tab1, gender.null)
45

```

```

46
47 # Age_base
48 age_base.tab1 <- lmer(age_base ~ cohort + (1|patient_id), data =
49 BARI_DATA )
50 age_base.null <- lmer(age_base ~ (1|patient_id), data = BARI_DATA)
51 anova(age_base.tab1, age_base.null)
52

```

```

53 # Disease duration
54 disease_duration_base_years.tab1 <- lmer(disease_duration_base_years ~
55 cohort + (1|patient_id), data = BARI_DATA )
56

```

```
1
2
3 disease_duration_base_years.null <- lmer(disease_duration_base_years ~
4 (1|patient_id), data = BARI_DATA)
5 anova(disease_duration_base_years.tab1,
6 disease_duration_base_years.null)
7
8
9
10 # CDAI raw
11 CDAI0_raw.tab1 <- lmer(CDAI0_raw ~ cohort + (1|patient_id), data =
12 BARI_DATA )
13 CDAI0_raw.null <- lmer(CDAI0_raw ~ (1|patient_id), data = BARI_DATA)
14 anova(CDAI0_raw.tab1, CDAI0_raw.null)
15
16
17 # CDAI (imputed)
18 CDAI0.tab1 <- lmer(CDAI0 ~ cohort + (1|patient_id), data = BARI_DATA )
19 CDAI0.null <- lmer(CDAI0 ~ (1|patient_id), data = BARI_DATA)
20 anova(CDAI0.tab1, CDAI0.null)
21
22
23
24 # obesity
25 obese_base.tab1 <- glmer(obese_base ~ cohort + (1|patient_id), data =
26 BARI_DATA , family = "binomial")
27 obese_base.null <- glmer(obese_base ~ (1|patient_id), data = BARI_DATA
28 , family = "binomial")
29 anova(obese_base.tab1, obese_base.null)
30
31
32
33 # smoker base - 1st level vs second level
34 smoker_base.tab1 <- glmer(smoker_base ~ cohort + (1|patient_id), data
35 = BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "FORMER_SMOKER")] ,
36 family = "binomial")
37 smoker_base.null <- glmer(smoker_base ~ (1|patient_id), data =
38 BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "FORMER_SMOKER")] ,
39 family = "binomial")
40 anova(smoker_base.tab1, smoker_base.null)
41
42
43
44 # smoker base - 2nd level versus third
45 smoker_base.tab1 <- glmer(smoker_base ~ cohort + (1|patient_id), data
46 = BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "NEVER_SMOKER")] ,
47 family = "binomial")
48 smoker_base.null <- glmer(smoker_base ~ (1|patient_id), data =
49 BARI_DATA[smoker_base %in% c("CURRENT_SMOKER", "NEVER_SMOKER")] ,
50 family = "binomial")
51 anova(smoker_base.tab1, smoker_base.null)
52
53
54
55 # seropositivity
56 seropositivity_base.tab1 <- glmer(seropositivity_base ~ cohort + (1|
```

```
1
2
3 patient_id), data = BARI_DATA , family = "binomial")
4 seropositivity_base.null <- glmer(seropositivity_base ~ (1|
5 patient_id), data = BARI_DATA , family = "binomial")
6 anova(seropositivity_base.tab1, seropositivity_base.null)
7
8
9
10 # Concomittant csDMARD
11 TC_with_csDMARD.tab1 <- glmer(TC_with_csDMARD ~ cohort + (1|
12 patient_id), data = BARI_DATA , family = "binomial")
13 TC_with_csDMARD.null <- glmer(TC_with_csDMARD ~ (1|patient_id), data =
14 BARI_DATA , family = "binomial")
15 anova(TC_with_csDMARD.tab1, TC_with_csDMARD.null)
16
17
18
19 # Line of therapy
20 line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
21 (1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "2nd")]
22 , family = "binomial")
23 line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
24 patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "2nd")] ,
25 family = "binomial")
26 anova(line_of_therapy.tab1, line_of_therapy.null)
27
28
29
30 # Line of therapy
31 line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
32 (1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "3rd")]
33 , family = "binomial")
34 line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
35 patient_id), data = BARI_DATA[line_of_therapy %in% c("1st", "3rd")] ,
36 family = "binomial")
37 anova(line_of_therapy.tab1, line_of_therapy.null)
38
39
40
41 # Line of therapy
42 line_of_therapy.tab1 <- glmer(as.factor(line_of_therapy) ~ cohort +
43 (1|patient_id), data = BARI_DATA[line_of_therapy %in% c("1st",
44 "4th_or_later")] , family = "binomial")
45 line_of_therapy.null <- glmer(as.factor(line_of_therapy) ~ (1|
46 patient_id), data = BARI_DATA[line_of_therapy %in% c("1st",
47 "4th_or_later")] , family = "binomial")
48 anova(line_of_therapy.tab1, line_of_therapy.null)
49
50
51
52 # N_prev_tsDMARD
53 N_prev_tsDMARD.tab1 <- glmer(as.factor(N_prev_tsDMARD) ~ cohort + (1|
54 patient_id), data = BARI_DATA , family = "binomial")
55 N_prev_tsDMARD.null <- glmer(as.factor(N_prev_tsDMARD) ~ (1|
56 patient_id), data = BARI_DATA , family = "binomial")
57
58
59
```

```

1
2
3 anova(N_prev_tsDMARD.tab1, N_prev_tsDMARD.null)
4
5
6 # Concomittant prednisone
7 PREDNISON_STEROID.tab1 <- glmer(PREDNISON_STEROID ~ cohort + (1|
8 patient_id), data = BARI_DATA , family = "binomial")
9 PREDNISON_STEROID.null <- glmer(PREDNISON_STEROID ~ (1|patient_id),
10 data = BARI_DATA , family = "binomial")
11 anova(PREDNISON_STEROID.tab1, PREDNISON_STEROID.null)
12
13
14
15 # Dose of PREDNISONE
16 PREDNISON_STEROID_dose.tab1 <- lmer(PREDNISON_STEROID_dose ~ cohort +
17 (1|patient_id), data = BARI_DATA )
18 PREDNISON_STEROID_dose.null <- lmer(PREDNISON_STEROID_dose ~ (1|
19 patient_id), data = BARI_DATA)
20 anova(PREDNISON_STEROID_dose.tab1, PREDNISON_STEROID_dose.null)
21
22
23 Other various computations for Table 1 --
24 uniqueN(BARI_DATA$patient_id)
25
26 mean(BARI_DATA[, time_on_drug])
27
28 median(BARI_DATA[cohort == "Bari" , time_on_drug])
29 median(BARI_DATA[cohort == "OMA" , time_on_drug])
30 median(BARI_DATA[cohort == "TNFi" , time_on_drug])
31
32 median(BARI_DATA[cohort == "OMA"]$time_on_drug)
33
34
35 mean(BARI_DATA[,disease_duration_base_years], na.rm = T)
36
37 table(is.na(BARI_DATA$CDAI0_raw), BARI_DATA$cohort) # number of
38 missing CDAI0_raw...
39 table(is.na(BARI_DATA$CDAI0), BARI_DATA$cohort) # number of missing
40 CDAI0... (after imputation)
41
42 table(is.na(BARI_DATA$CDAI12_raw), BARI_DATA$cohort) # number of
43 missing CDAI12_raw...
44 table(is.na(BARI_DATA$CDAI12), BARI_DATA$cohort) # number of missing
45 CDAI12 after imputation
46
47
48 hist(BARI_DATA$CDAI0_raw)
49 hist(BARI_DATA$CDAI0)
50
51 summary(BARI_DATA[cohort=="BARI", c("gender",
52 "age_base","disease_duration_base_years", "CDAI0_raw", "CDAI0",
53 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
54 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
55 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
56
57
58
59
60

```

```
1
2
3 "initiation_year", "time_on_drug", "HAQ_score_base")) # to see NA
4 values for all variables
5
```

```
6
7 summary(BARI_DATA[cohort=="TNFi", c("gender",
8 "age_base", "disease_duration_base_years", "CDAI0_raw", "CDAI0",
9 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
10 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
11 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
12 "initiation_year", "time_on_drug", "HAQ_score_base")) # to see NA
13 values for all variables
14
```

```
15
16 summary(BARI_DATA[cohort=="OMA", c("gender",
17 "age_base", "disease_duration_base_years", "CDAI0_raw", "CDAI0",
18 "obese_base", "smoker_base", "seropositivity_base", "time_on_drug365",
19 "TC_with_csDMARD", "line_of_therapy", "N_prev_tsDMARD",
20 "PREDNISON_STEROID", "PREDNISON_STEROID_dose", "dose",
21 "initiation_year", "time_on_drug", "HAQ_score_base")) # to see NA
22 values for all variables
23
```

```
24 table(is.na(BARI_DATA$disease_duration_base_years), BARI_DATA$cohort)
25 # number of missing disease duration...
26
```

```
27 table(is.na(BARI_DATA$age_base), BARI_DATA$cohort) # number of steroid
28 doses missing
29
```

```
30 table(is.na(BARI_DATA$PREDNISON_STEROID_dose), BARI_DATA$cohort) #
31 number of missing baseline steroids
32
```

33 Imputation using MICE -- BARI vs TNFi et OMA bDMARDs --

34
35 Common imputation step with all data

```
36
37 BARI <- BARI_DATA[,c("TC_id", "patient_id", "stop_DMARD",
38 "stop_reasons", "age_base", "concomitant_csDMARD",
39 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
40 "CDAI0", "disease_duration_base_years", "time_on_drug", "bmi_base",
41 "smoker_base", "line_of_therapy", "obese_base", "gender", "cohort",
42 "adverse_event_reported", "seropositivity_base")] # choose variables
43 of interest
44
```

```
45
46 BARI$smoker_base <- as.factor(BARI$smoker_base) # put labels as factor
47 BARI$line_of_therapy <- as.factor(BARI$line_of_therapy)
48 BARI$gender <- as.factor(BARI$gender)
49 BARI$concomitant_csDMARD <- as.factor(BARI$concomitant_csDMARD)
50 BARI$PREDNISON_STEROID <- as.factor(BARI$PREDNISON_STEROID)
51 BARI$cohort <- as.factor(BARI$cohort)
52
```

53
54 # Imputation

```
55
56 if(!file.exists("./2_cached_files/imputed_data")){ # to avoid re-
57
58
59
```

```

1
2
3   computing if already done
4     imputed_data <- mice(BARI, m=50, method="pmm", maxit=25, seed=500)
5     save(imputed_data, file = "./2_cached_files/imputed_data")
6   } else {
7     load("./2_cached_files/imputed_data")
8   }
9
10  # Subsettings
11
12  BARI1 <- BARI[cohort %in% c("BARI", "TNFi")] # creating subset for
13  BARI vs TNFi comparaisn
14  BARI1[,cohort := as.factor(as.character(cohort))]
15
16
17  imputed_data1 <- complete(imputed_data,"long", include=T) # to put in
18  long format and categorize variables
19  imputed_data1 <- imputed_data1[imputed_data1$cohort %in% c("BARI",
20  "TNFi"),] # to keep only BARI and TNFi rows
21  imputed_data1$cohort <- as.factor(as.character(imputed_data1$cohort))
22  imputed_data1 <- as.mids(imputed_data1) # re concateneting in previous
23  format, to use fit.mult.impute
24
25
26
27  BARI2 <- BARI[cohort %in% c("BARI", "OMA")] # creating subset for BARI
28  vs OMA comparaisn
29  BARI2[,cohort := as.factor(as.character(cohort))]
30
31
32  imputed_data2 <- complete(imputed_data,"long", include=T) # to put in
33  long format and categorize variables
34  imputed_data2 <- imputed_data2[imputed_data2$cohort %in% c("BARI",
35  "OMA"),] # to keep only BARI and OMA rows
36  imputed_data2$cohort <- as.factor(as.character(imputed_data2$cohort))
37  imputed_data2 <- as.mids(imputed_data2) # re concateneting in previous
38  format, to use fit.mult.impute
39
40

```

1. [1] SURVIVAL ANALYSIS (drug discontinuation)

Exploration

```

44 table(BARI_DATA$cohort, BARI_DATA$stop_DMARD)
45 table(BARI_DATA$cohort, BARI_DATA$stop_reasons)
46

```

Checking adequacy of COX models --

For BARI vs TNFi

```

51 # categorization for linearity checking
52 test1 <- complete(imputed_data1,"long", include=T) # to put in long
53 format and categorize variables
54
55
56
57
58
59

```



```
1
2
3 test1$agecat <- cut(test1$age_base, 4)
4 test1$bmicat <- cut(test1$bmi_base, 4)
5 test1$cdaicat <- cut(test1$CDAI0, 4)
6 test1$duracat <- cut(test1$disease_duration_base_years, 4)
7
8
9 test1 <- as.mids(test1, .imp=1, .id=2) # re concatenating in previous
10 format, to use fit.mult.impute
11
12 # linearity checking
13
14 BARI1.adj.mi.test <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
15 as.factor(cohort)+
16                                     I(agecat)+
17                                     I(bmicat)+
18                                     TC_with_csDMARD+
19                                     PREDNISON_STEROID+
20                                     I(cdaicat)+
21                                     I(duracat)+
22                                     C(smoker_base, base=3)+
23                                     line_of_therapy+
24                                     gender+
25                                     seropositivity_base+
26                                     cluster(patient_id),
27                                     fitter = coxph, xtrans = test1,
28                                     data = BARI1)
29
30
31 summary(BARI1.adj.mi.test)
32 rm(BARI1.adj.mi.test)
33
34 # Log-linearity of coefficients ?
35
36 # Coefs age are between 0.15 and 0.25, let's assume it's ok
37 # Hum bmi coefs are not so log-linear, rather close to 0
38 # For CDAI also
39 # Looks ok for disease_duration_base_years.
40
41 # --> Let's keep all variable in the continuous format
42
43 # Proportionality test of hazards on raw data
44
45
46 test1ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
47 as.factor(cohort)+
48                                     cluster(patient_id),
49                                     data= BARI1)
50
51 cox.zph(test1ph) # it's ok
52
53 # Hazard proportionality test on imputed data sets
54
55 test1 <- complete(test1,"long",include=T) # To reset the charges to
56
57
58
59
60
```

```
1
2
3 long format
4 test1 <- test1[test1$.imp==1 | test1$.imp==2 | test1$.imp==3 |
5 test1$.imp==4 | test1$.imp==5 ,] # To select only 5 data sets
6
7 test1ph.adj.mi <- coxph(Surv(time = time_on_drug, event = stop_DMARD)
8 ~ as.factor(cohort)+
9
10         I(age_base/10)+
11         bmi_base+
12         TC_with_csDMARD+
13         PREDNISON_STEROID+
14         CDAI0+
15         I(disease_duration_base_years/10)+
16         C(smoker_base, base=3)+
17         line_of_therapy+
18         gender+
19         seropositivity_base+
20         cluster(patient_id),
21         data = test1)
22
23
24 cox.zph(test1ph.adj.mi)
25
26 schonfeldall <- cox.zph(test1ph.adj.mi) # Test cox.zph may not be ok,
27 but it's because of the multiple imputation (often the case with a lot
28 of data)
29 for (i in 1:11){
30     plot(schonfeldall[i]) # so we should go to a visual testing --> ok
31 }
32
33 rm(schonfeldall, test1ph.adj.mi, test1)
34
35 For BARI vs OMA
36
37 # categorization for linearity checking
38 test2 <- complete(imputed_data2,"long", include=T) # to put in long
39 format and categorize variables
40
41 test2$agecat <- cut(test2$age_base, 4)
42 test2$bmicat <- cut(test2$bmi_base, 4)
43 test2$cdai0cat <- cut(test2$CDAI0, 4)
44 test2$duracat <- cut(test2$disease_duration_base_years, 4)
45
46
47 test2 <- as.mids(test2, .imp=1, .id=2) # re concatenating in previous
48 format, to use fit.mult.impute
49
50 # linearity checking
51
52 BARI2.adj.mi.test <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
53 as.factor(cohort)+
54
55         I(agecat)+
56         I(bmicat)+
57
58
59
60
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

TC_with_csDMARD+
PRÉDNISÓN_STERÓID+
I(cdaicat)+
I(duracat)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base+
cluster(patient_id),
fitter = coxph, xtrans = test2,
data = BARI2)

summary(BARI2.adj.mi.test)
rm(BARI2.adj.mi.test)

# Log-linearity of coefficients ?

# Coefs age are around -0.4, let's assume it's ok
# Hum bmi coefs are discussable
# For CDAI it's ok
# More or less ok for disease_duration_base_years.

# --> Let's keep all variable in the continuous format

# Proportionality test of hazards on raw data

test2ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
as.factor(cohort)+
cluster(patient_id),
data= BARI2)
cox.zph(test2ph) # it's ok

# Hazard proportionality test on imputed data sets

test2 <- complete(test2,"long",include=T) # To put imputed data in
ling format
test2 <- test2[test2$.imp==1 | test2$.imp==2 | test2$.imp==3 |
test2$.imp==4 | test2$.imp==5 ,] # To select only 5 datasets

test2ph.adj.mi <- coxph(Surv(time = time_on_drug, event = stop_DMARD)
~ as.factor(cohort)+
I(age_base/10)+
bmi_base+
TC_with_csDMARD+
PRÉDNISÓN_STERÓID+
CDAI0+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+

```

```

1
2
3           line_of_therapy+
4           gender+
5           seropositivity_base+
6           cluster(patient_id),
7           data=test2)
8
9
10      cox.zph(test2ph.adj.mi)
11
12      schonfeldall <- cox.zph(test2ph.adj.mi) # Test cox.zph may not be ok,
13      but it's because of the multiple imputation (often the case with a lot
14      of data)
15      for (i in 1:11){
16          plot(schonfeldall[i]) # so we should go to a visual testing --> ok
17      }
18
19      rm(schonfeldall, test2ph.adj.mi, test2)

```

BARI vs TNFi --

COX model

Final Cox Model

```

27      BARI1.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
28      cohort +
29
30          I(age_base/10)+
31          bmi_base+
32          TC_with_csDMARD+
33          PREDNISON_STEROID+
34          I(CDAI0/10)+
35          I(disease_duration_base_years/10)+
36          C(smoker_base, base=3)+
37          line_of_therapy+
38          gender+
39          seropositivity_base+
40          cluster(patient_id),
41          fitter = coxph, xtrans = imputed_data1,
42          data = BARI1)
43
44      summary(BARI1.adj.mi)

```

Creation of HR table and p-values

```

48      ploufrows <- names(BARI1.adj.mi$coefficients)
49      ploufcols <- c("HR", "95%CI", "p")
50      coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
51      length(ploufcols))
52      rownames(coxtable) <- ploufrows
53      colnames(coxtable) <- ploufcols
54      plouf <- summary(BARI1.adj.mi)

```

```

1
2
3 for(row in ploufrows)
4 {
5   coxtable[row,"HR"] <-
6   formattable(plouf$coefficients[row,"exp(coef)"])
7   coxtable[row,"95%CI"] <-
8   paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",
9   formattable(plouf$conf.int[row,"upper .95"]))
10  coxtable[row,"p"] <- writepvalue(plouf$coefficients[row,"Pr(>|
11  z|)"])
12 }
13
14
15 write.xlsx(coxtable, file="./3_clean_output/BARI vs TNFi HR.xlsx") #
16 saving excel file
17
18 Forest plot
19 meanall <- summary(BARI1.adj.mi)$coefficients[1:14,"exp(coef)"]
20 lowerall <- summary(BARI1.adj.mi)$conf.int[1:14,"lower .95"]
21 upperall <- summary(BARI1.adj.mi)$conf.int[1:14,"upper .95"]
22 textall <- c("TNFi (vs BARI)", "Age (decades)", "BMI", "Concomitant
23 csDMARD", "Concomitant glucocorticoid", "CDAI score (10 pts)",
24 "Disease duration (decades)", "Current smoker (vs non-smoker)",
25 "Former smoker (vs non-smoker)", "2nd line therapy (vs 1st)", "3rd
26 line therapy (vs 1st)", "4th or later line (vs 1st)", "Female gender",
27 "Seropositivity (RF or ACPA)")
28
29
30 dfall <- data.frame(textall, meanall, lowerall, upperall)
31 dfall$textall <- factor(dfall$textall,
32   levels = textall)
33
34 HR_plot_1 <- ggplot(data=dfall, aes(x=textall, y= meanall, ymin =
35 lowerall, ymax = upperall))+
36
37   geom_pointrange(size=0.5)+
38   geom_errorbar(aes(ymin=lowerall, ymax=upperall),width=0.5)+
39   geom_hline(yintercept =1, linetype=2)+
40
41   xlab('')+ ylab(" ")
42   ggtitle("BARI vs TNFi")+
43
44   scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
45 +
46   facet_wrap(~textall,nrow=16, strip.position= "right", scales =
47 "free_y") +
48
49
50   theme_pubclean()+
51   theme(strip.text.y = element_blank(),
52     strip.background = element_blank(),
53     axis.line.x = element_line(size = 0.5),
54     axis.text = element_text(face = "bold", colour = "black"),
55     legend.position="bottom", plot.margin =
56
57
58
59
60

```

```
1
2
3     unit(c(1,3,2,1),"lines"))+
4
5         coord_flip()
6
7 HR_plot_1
8
9
10 # adding some manual annotation
11 grid.text("Improves drug maintenance", x = unit(0.3, "npc"), y =
12 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
13 grid.text("Reduces drug maintenance", x = unit(0.87, "npc"), y =
14 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
15
```

Non-adjusted Kaplan-Meier curves

based on mini-tutorial found on datacamp.com/community/tutorials/survival-analysis-R

BARI vs TNFi

```
21 surv_object1 <- Surv(time = BARI1$time_on_drug, event =
22 BARI1$stop_DMARD) # indicate time on drug and stop variable
23 summary(coxph(surv_object1 ~ cohort, data=BARI1))
24 fit1 <- survfit(surv_object1 ~ cohort, data = BARI1) # this function
25 creates the data for Kaplan Meyer
26 fit1
27
28 survplot_1 <- ggsurvplot(fit1, data = BARI1, # plot
29 pval = T,
30 pval.method = TRUE,
31 legend.title = "Groups :",
32 legend.labs = c("Baricitinib", "TNFi"),
33 xlab = "Time (days)",
34 xlim = c(0, 700),
35 censor = FALSE,
36 title = "Non-adjusted model of drug discontinuation by type
37 of treatment",
38 surv.median.line = "v",
39 linetype = 1,
40 size = 1.5,
41 ggtheme = theme_minimal(),
42 #palette = c("grey78", "grey10"),
43 palette = c("red2", "green3"), # specify colors
44 risk.table = T)
45
46 survplot_1
47 summary(fit1, times = 365)
48 summary(fit1, times = 730)
49
```

Saving the plot curv object for Lilly

```
50
51
52 plot_BARI_vs_TNFi_data <- survplot_1$data.survplot
53 write.xlsx(plot_BARI_vs_TNFi_data, file =
54 "./3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_non_adjust
55 ed.xlsx", row.names = F)
56
57
58
59
```

Home-made attempt to obtain adjusted curves based on imputed data

```
1
2
3
4 dummy_cox_impute1 <- mice::complete(imputed_data1, "long", include =
5 T)
6 dummy_cox_impute1 <- dummy_cox_impute1[dummy_cox_impute1$.imp != 0,]
7
8
9 BARI_fit1 <- survfit(coxph(Surv(time = time_on_drug, event =
10 stop_DMARD) ~ cohort+
11 I(age_base/10)+
12 bmi_base+
13 TC_with_csDMARD+
14 PREDNISON_STEROID+
15 CDAI0+
16 I(disease_duration_base_years/10)+
17 C(smoker_base, base=3)+
18 line_of_therapy+
19 gender+
20 seropositivity_base+
21 cluster(patient_id)+
22 strata(cohort),dummy_cox_impute1), data =
23 dummy_cox_impute1)
24
25
26
27 survplot_1_adj <- ggsurvplot(BARI_fit1, data = dummy_cox_impute1,
28 variable = "cohort",
29 xlab = "Time (days)",
30 title = "Multivariable Cox model of drug discontinuation by
31 type of treatment - BARI vs TNFi",
32 legend.title = "Groups :",
33 legend.labs = c("Baricitinib", "TNFi"),
34 censor = FALSE,
35 xlim = c(0, 700),
36 surv.median.line = "v",
37 linetype = 1,
38 size = 1.5,
39 ggtheme = theme_minimal(),
40 # palette = c("grey78", "grey10")
41 palette = c("red2", "green3") # to change colors
42 )
43
44
45 # adding some legends
46 survplot_1_adj <- survplot_1_adj +
47 labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
48 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
49 smoking status, line of therapy, gender, seropositivity")
50
51
52 survplot_1_adj
53 # summary(BARI_fit1) # to see detailed surv probabilities at given
54 timepoints
55
56
57
58
59
60
```

```
1
2
3 summary(BARI_fit1, times = 365)
4 summary(BARI_fit1, times = 730)
5
```

```
6 Saving the plot curv object for Lilly
7
```

```
8 plot_BARI_vs_TNFi_data_adj <- survplot_1_adj$data.survplot
9 write.xlsx(plot_BARI_vs_TNFi_data_adj, file =
10 ".\3_clean_output\Lilly_curves_excel/plot_BARI_vs_TNFi_data_adj.xlsx",
11 row.names = F)
12
```

Sensitivity analysis with package RiskRegression (AIPTW)

```
14 # Rappel: imputed_data1 = BARI vs TNFi
15 #           imputed_data2 = BARI vs OMA
16 library(riskRegression)
17
```

```
18 # I select only one imputed dataset. Would be even better to find a
19 way to pool/average the results from the 50 imputed datasets, but it
20 does not seem doable by default
21 test.data <- complete(imputed_data1, 1)
22
23
```

```
24 # First, we specify the treatment model (propensity score model)
25 # Logistic regression where the treatment group is the dependent
26 variable.
27
```

```
28 m.treatment <- glm(cohort~I(age_base/10)+
29                    bmi_base+
30                    TC_with_csDMARD+
31                    PREDNISON_STEROID+
32                    I(CDAI0/10)+
33                    I(disease_duration_base_years/10)+
34                    C(smoker_base, base=3)+
35                    line_of_therapy+
36                    gender+
37                    seropositivity_base,
38                    data = test.data, family =
39                    "binomial" )
40
41
```

```
42 # Then we specify both the "event model" and the "censoring model".
43 Both are cox model
44
```

```
45 m.event <- coxph(Surv(time_on_drug, stop_DMARD) ~ cohort+
46                 I(age_base/10)+
47                 bmi_base+
48                 TC_with_csDMARD+
49                 PREDNISON_STEROID+
50                 I(CDAI0/10)+
51                 I(disease_duration_base_years/10)+
52                 C(smoker_base, base=3)+
53                 line_of_therapy+
54                 gender+
55                 )
56
57
58
59
```



```
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

```
seropositivity_base,
data = test.data, x = TRUE, y =
TRUE)

m.censor <- coxph(Surv(time_on_drug, stop_DMARD==0) ~ cohort +
                  I(age_base/10)+
                  bmi_base+
                  TC_with_csDMARD+
                  PREDNISON_STEROID+
                  I(CDAI0/10)+
                  I(disease_duration_base_years/10)+
                  C(smoker_base, base=3)+
                  line_of_therapy+
                  gender+
                  seropositivity_base
                  , x =TRUE, y = TRUE,
                  data = test.data)

# And we measure the average treatment effect using function "ate",
specifying the time at which we want to compute the ATE

out <- ate(event = m.event ,
           treatment = m.treatment,
           censor = m.censor,
           data = test.data,
           cause = 1,
           estimator = "AIPTW",
           times = seq(from = 0, to = 500, by = 5))

dt.out <- as.data.table(out)

Diagnostics asked by Lily statistician

library(cobalt)

# First, the distribution of propensity scores
test.data$pscores <- m.treatment$fitted.values
test.data %>% setDT()

pscore_plot <- ggplot(test.data, aes(x = pscores, color = cohort, fill
= cohort)) +
  geom_density(alpha = .47) +
  xlab("Estimated Probability of being assigned BARI") +
  ylab("Density") +
  theme_minimal()+
  theme(axis.ticks.y = element_blank(),
        panel.grid.minor = element_blank(),
        legend.title = element_blank(),
        text = element_text(size = 16),
```

```
1
2
3     axis.title.x = element_text(hjust = 0.2, size = 16))
4 pscore_plot # overlap
5
6
7
8 ## Computing the weights
9 test.data$weights <- ifelse(test.data$cohort == "TNFi",
10 1/test.data$pscores, 1/(1-test.data$pscores))
11
12 # Selecting only our covariates of interest (the ones in the ps model)
13 COVS <- subset(test.data, select = c(cohort,age_base,
14     bmi_base,
15     TC_with_csDMARD,
16     PREDNISOLONE_STEROID,
17     CDAI0,
18     disease_duration_base_years,
19     smoker_base,
20     line_of_therapy,
21     gender,
22     seropositivity_base))
23
24
25 # To get the SMD & variance ratios before/after weighting
26 # bal.tab(COVS, treat = test.data$cohort, thresholds = 0.1)
27 # bal.tab(COVS, treat = test.data$cohort, weights = test.data$weights,
28 thresholds = 0.1)
29 # bal.tab(COVS, treat = test.data$cohort, v.threshold = 2)
30 # bal.tab(COVS, treat = test.data$cohort, weights = test.data$weights,
31 v.threshold = 2)
32
33
34
35 # But plotting is clearer:
36 love.plot(COVS, treat = test.data$cohort, weights = test.data$weights,
37 stats = c("mean.diffs"), thresholds = c(m = .1), var.order =
38 "adjusted")
39
40 # We can also plot variance ratios for continuous variables
41 love.plot(COVS, treat = test.data$cohort, weights = test.data$weights,
42 stats = c("variance.ratios"))
43
44 # propensity scores enhance the balance overall, except for the CDAI0.
45 However, this is the reason we use the AIPTW. The remaining imbalance
46 is accounted for by the outcome model (outcome model is the cox
47 regression), and the misspecification of the outcome model is
48 mitigated by the balancing done by propensity score.
49
50
51 First plot to get the difference in average treatment effect in percentage
52
53 plot.ate.diff <- ggplot(dt.out[type == "meanRisk"], aes(x = time,
54 group = level))+
55   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
56 0.4)+
57
58
59
```

```

1
2
3   geom_line(aes(y = estimate, color = level), size = 1)+
4   #geom_vline(xintercept = 90)+
5
6   scale_colour_manual(values = c("lightblue","darkseagreen3"))+
7   scale_fill_manual(values = c("lightblue","darkseagreen3"))+
8   theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
9   legend.position="top" )+
10  scale_x_continuous(breaks=seq(0,500,50)) + scale_y_continuous(labels
11  = scales::percent)+
12
13  xlab("Days since intiation of treatment")+
14  ylab("Absolute Risk of treatment discontinuation (%)")+
15  labs(colour="Groups:", fill = "Groups:")+
16  labs(group = "Groups:")+
17  theme_bw(base_size = 14)+
18  theme(axis.title.x = element_text(margin = margin(t = .3,unit =
19  "cm")),
20  axis.title.y = element_text(margin = margin(r = .3,unit =
21  "cm")))
22
23
24 plot.ate.diff

```

Second plot to get the ratio in average treatment effect

```

28 plot.ate.ratio <- ggplot(dt.out[type == "ratioRisk"], aes(x = time,
29 group = level))+
30   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
31   0.4)+
32   geom_line(aes(y = estimate, color = level), size = 2)+
33
34   theme_minimal()+
35   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
36   scale_x_continuous(breaks=seq(0,500,50))+
37   scale_y_continuous(limits = c(0.9,4.5))+
38
39   xlab("Days since intiation of treatment")+
40   ylab("Ratio in Average Treatment Effect")+
41   labs(colour="treatment", fill = "treatment")
42
43
44 plot.ate.ratio

```

We can also consider the AIPTW estimate at a specific time point. For example at 365-day.

```

48 r.one <- dt.out[type == "diffRisk" & time == 365, .
49 (estimate,lower,upper,p.value)]
50 r.two <- dt.out[type == "ratioRisk" & time == 365, .
51 (estimate,lower,upper,p.value)]
52

```

```

54 ploufrows <- c("Difference in average treatment effect","Ratio in
55 average treatment effect")
56 ploufcols <- c("Estimate","95%CI","p")
57
58
59
60

```

```

1
2
3 table <- matrix(data = NA, nrow = length(ploufrows), ncol =
4 length(ploufcols))
5 rownames(table) <- ploufrows
6 colnames(table) <- ploufcols
7
8
9 library(formattable)
10 table[1,"Estimate"] <- paste0(formattable(r.one$estimate*100),"%")
11 table[1,"95%CI"] <-
12 paste0(formattable(r.one$lower),"-",formattable(r.one$upper))
13 table[1,"p"] <- writepvalue(r.one$p.value)
14 table[2,"Estimate"] <- paste0(r.two$estimate)
15 table[2,"95%CI"] <- paste0(r.two$lower,"-",r.two$upper)
16 table[2,"p"] <- writepvalue(r.two$p.value)
17
18 table
19
20 # Interpretation: If every patient had received BARI, the 365-day risk
21 of treatment discontinuation would have been 19.34% (points) lower
22 compared to when every patient had received TNFi.

```

BARI vs OMA --

COX model

```

28 BARI2.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
29 cohort+
30
31                               I(age_base/10)+
32                               bmi_base+
33                               TC_with_csDMARD+
34                               PREDNISON_STEROID+
35                               I(CDAI0/10)+
36                               I(disease_duration_base_years/10)+
37                               C(smoker_base, base=3)+
38                               line_of_therapy+
39                               gender+
40                               seropositivity_base+
41                               cluster(patient_id),
42                               fitter = coxph, xtrans = imputed_data2,
43                               data = BARI2)
44
45 summary(BARI2.adj.mi)

```

Creation of HR table and p-values (denis)

```

48 ploufrows <- names(BARI2.adj.mi$coefficients)
49 ploufcols <- c("HR","95%CI","p")
50 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
51 length(ploufcols))
52 rownames(coxtable) <- ploufrows
53 colnames(coxtable) <- ploufcols
54 plouf <- summary(BARI2.adj.mi)

```

```

1
2
3 for(row in ploufrows)
4 {
5   coxtable[row,"HR"] <-
6   formattable(plouf$coefficients[row,"exp(coef)"])
7   coxtable[row,"95%CI"] <-
8   paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",formattable(pl
9   ouf$conf.int[row,"upper .95"]))
10   coxtable[row,"p"] <- writepvalue(plouf$coefficients[row,"Pr(>|
11   z|)"])
12 }
13
14
15 write.xlsx(coxtable, file="./3_clean_output/BARI vs OMA HR.xlsx") #
16 saving excel file
17

```

Forest plot

```

18
19 meanall <- summary(BARI2.adj.mi)$coefficients[1:14,"exp(coef)"]
20 lowerall <- summary(BARI2.adj.mi)$conf.int[1:14,"lower .95"]
21 upperall <- summary(BARI2.adj.mi)$conf.int[1:14,"upper .95"]
22 textall <- c("OMA (vs BARI)", "Age (decades)", "BMI", "Concomitant
23 csDMARD", "Concomitant glucocorticoid", "CDAI score (10 pts)",
24 "Disease duration (decades)", "Current smoker (vs non-smoker)",
25 "Former smoker (vs non-smoker)", "2nd line therapy (vs
26 1st)", "3rd line therapy (vs 1st)", "4th or later line (vs 1st)",
27 "Female gender", "Seropositivity (RF or ACPA)")
28
29
30 dfall <- data.frame(textall, meanall, lowerall, upperall)
31 dfall$textall <- factor(dfall$textall,
32 levels = textall)
33
34 HR_plot_2 <- ggplot(data=dfall, aes(x=textall, y= meanall, ymin =
35 lowerall, ymax = upperall))+
36
37   geom_pointrange(size=0.5)+
38   geom_errorbar(aes(ymin=lowerall, ymax=upperall),width=0.5)+
39   geom_hline(yintercept =1, linetype=2)+
40
41   xlab('')+ ylab(" ")
42   ggtitle("BARI vs OMA")+
43
44   scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
45 +
46   facet_wrap(~textall,nrow=16, strip.position= "right", scales =
47 "free_y") +
48
49
50   theme_pubclean()+
51   theme(strip.text.y = element_blank(),
52         strip.background = element_blank(),
53         axis.line.x = element_line(size = 0.5),
54         axis.text = element_text(face = "bold", colour = "black"),
55         legend.position="bottom", plot.margin =
56
57
58
59
60

```

```
1
2
3     unit(c(1,3,2,1),"lines"))+
4
5         coord_flip()
6
7 HR_plot_2
8
9
10 # adding some manual annotation
11 grid.text("Improves drug maintenance", x = unit(0.3, "npc"), y =
12 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
13 grid.text("Reduces drug maintenance", x = unit(0.87, "npc"), y =
14 unit(0.05, "npc"), gp = gpar(fontface = "bold"))
15
```

Non-adjusted Kaplan-Meier curves

based on mini-tutorial found on datacamp.com/community/tutorials/survival-analysis-R)

BARI vs OMA

```
21 surv_object2 <- Surv(time = BARI2$time_on_drug, event =
22 BARI2$stop_DMARD)
23 fit2 <- survfit(surv_object2 ~ cohort, data = BARI2) # this function
24 creates the data for Kaplan Meyer
25 survplot_2 <- ggsurvplot(fit2, data = BARI2, # plot
26   pval = T,
27   pval.method = TRUE,
28   legend.title = "Groups :",
29   legend.labs = c("Baricitinib", "OMA"),
30   xlab = "Time (days)",
31   xlim = c(0, 700),
32   censor = FALSE,
33   title = "Non-adjusted model of drug discontinuation by type
34 of treatment",
35   surv.median.line = "v",
36   linetype = 1,
37   size = 1.5,
38   ggtheme = theme_minimal(),
39   #palette = c("grey78", "grey50"),
40   palette = c("red2", "blue3"), # to put colors
41   risk.table = T)
42
43 survplot_2
44 summary(fit2, times = 365)
45 summary(fit2, times = 730)
```

Saving the plot curv object for Lilly

```
46
47
48 plot_BARI_vs_OMA_data <- survplot_2$data.survplot
49 write.xlsx(plot_BARI_vs_OMA_data, file =
50 ".\\3_clean_output\\Lilly_curves_excel\\plot_BARI_vs_OMA_data_non_adjuste
51 d.xlsx", row.names = F)
52
53
54
55
56
57
58
59
```

Home-made attempt to obtain adjusted cuves based on imputed data :

```

1
2
3
4 dummy_cox_impute2 <- mice::complete(imputed_data2, "long", include =
5 T)
6 dummy_cox_impute2 <- dummy_cox_impute2[dummy_cox_impute2$.imp != 0,]
7
8
9 BARI_fit2 <- survfit(coxph(Surv(time = time_on_drug, event =
10 stop_DMARD) ~ cohort+
11           I(age_base/10)+
12           bmi_base+
13           TC_with_csDMARD+
14           PREDNISON_STEROID+
15           CDAI0+
16           I(disease_duration_base_years/10)+
17           C(smoker_base, base=3)+
18           line_of_therapy+
19           gender+
20           seropositivity_base+
21           cluster(patient_id)+
22           strata(cohort),dummy_cox_impute2), data =
23 dummy_cox_impute2)
24
25
26 survplot_2_adj <- ggsurvplot(BARI_fit2, data = dummy_cox_impute2,
27 variable = "cohort",
28       xlab = "Time (days)",
29       title = "Multivariable Cox model of drug discontinuation by
30 type of treatment - BARI vs OMA",
31       legend.title = "Groups :",
32       legend.labs = c("Baricitinib", "OMA bDMARDs"),
33       censor = FALSE,
34       xlim = c(0, 700),
35       surv.median.line = "v",
36       linetype = 1,
37       size = 1.5,
38       ggtheme = theme_minimal(),
39       # palette = c("grey78", "grey50")
40       palette = c("red2", "blue3") # to change colors
41     )
42
43
44 # adding some legends
45 survplot_2_adj <- survplot_2_adj +
46   labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
47 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
48 smoking status, line of therapy, gender, seropositivity")
49
50
51 survplot_2_adj
52 summary(BARI_fit2, times = 365) # to see detailed surv probabilities
53 at given timepoints
54 summary(BARI_fit2, times = 730)
55
56 Saving the plot curv object for Lilly
57
58
59
60

```

```

1
2
3 plot_BARI_vs_OMA_data_adj <- survplot_2_adj$data.survplot
4 write.xlsx(plot_BARI_vs_OMA_data_adj, file =
5 ".\3_clean_output\Lilly_curves_excel\plot_BARI_vs_OMA_data_adj.xlsx",
6 row.names = F)
7

```

Sensitivity analysis with package RiskRegression (AIPTW)

```

8
9 # I select only one imputed dataset. Would be good to find a way to
10 pool the results from the 50 datasets imputed.
11

```

```

12 test.data2 <- complete(imputed_data2,1)
13

```

```

14 # First, we specify the treatment model (propensity score model)
15 # Logistic regression where the treatment group is the dependent
16 variable.
17

```

```

18 m.treatment2 <- glm(cohort~I(age_base/10)+
19                    bmi_base+
20                    TC_with_csDMARD+
21                    PREDNISON_STEROID+
22                    I(CDAI0/10)+
23                    I(disease_duration_base_years/10)+
24                    C(smoker_base, base=3)+
25                    line_of_therapy+
26                    gender+
27                    seropositivity_base,
28                    data = test.data2, family =
29                    "binomial" )
30

```

```

31
32 # Then we specify both the "event model" and the "censoring model".
33 Both are cox model
34

```

```

35 m.event2 <- coxph(Surv(time_on_drug, stop_DMARD) ~ cohort+
36                  I(age_base/10)+
37                  bmi_base+
38                  TC_with_csDMARD+
39                  PREDNISON_STEROID+
40                  I(CDAI0/10)+
41                  I(disease_duration_base_years/10)+
42                  C(smoker_base, base=3)+
43                  line_of_therapy+
44                  gender+
45                  seropositivity_base,
46                  data = test.data2, x = TRUE, y =
47                  TRUE)
48

```

```

49
50 m.censor2 <- coxph(Surv(time_on_drug,stop_DMARD==0) ~ cohort +
51                  I(age_base/10)+
52                  bmi_base+
53                  TC_with_csDMARD+
54                  PREDNISON_STEROID+
55                  I(CDAI0/10)+
56

```



```
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

```
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base
, x =TRUE, y = TRUE,
data = test.data2)

# And we measure the average treatment effect using function "ate",
specifying the times at which we want to compute the ATE

out2 <- ate(event = m.event2 ,
            treatment = m.treatment2,
            censor = m.censor2,
            data = test.data2,
            cause = 1,
            estimator = "AIPTW",
            times = seq(from = 0, to = 500, by = 5))

dt.out2 <- as.data.table(out2)

Diagnostics asked by Lily statistician

library(cobalt)

# First, the distribution of propensity scores
test.data2$pscores <- m.treatment2$fitted.values
test.data2 %>% setDT()

pscore_plot2 <- ggplot(test.data2, aes(x = pscores, color = cohort,
fill = cohort)) +
  geom_density(alpha = .47) +
  xlab("Estimated Probability of being assigned BARI") +
  ylab("Density") +
  theme_minimal()+
  theme(axis.ticks.y=element_blank(),
        panel.grid.minor=element_blank(),
        legend.title=element_blank(),
        text = element_text(size = 16),
        axis.title.x =element_text(hjust = 0.2, size = 16))
pscore_plot2
# Good overlap

## Computing the weights
test.data2$weights <- ifelse(test.data2$cohort == "OMA",
1/test.data2$pscores, 1/(1-test.data2$pscores))

# Selecting only our covariates of interest (the ones in the ps model)
COVS_2 <- subset(test.data2, select = c(cohort,age_base,
bmi_base,
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

TC_with_csDMARD,
PRÉDNISÓN_STEROID,
CDAI0,
disease_duration_base_years,
smoker_base,
line_of_therapy,
gender,
seropositivity_base))

# To get the SMD & variance ratios before/after weighting
# bal.tab(COVS_2, treat = test.data2$cohort, thresholds = 0.1)
# bal.tab(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, thresholds = 0.1)
# bal.tab(COVS_2, treat = test.data2$cohort, v.threshold = 2)
# bal.tab(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, v.threshold = 2)
#

#But plotting it is better:
love.plot(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
var.order = "adjusted")

# We can also plot variance ratios for continuous variables
love.plot(COVS_2, treat = test.data2$cohort, weights =
test.data2$weights,stats = c("variance.ratios"))

# propensity scores enhance the balance overall, except for the CDAI0.
However, this is the reason we use the AIPTW. The remaining imbalance
is accounted for by the outcome model (outcome model is the cox
regression), and the misspecification of the outcome model is
mitigated by the balancing done by propensity score.

First plot to get the difference in average treatment effect in percentage

plot.ate.diff2 <- ggplot(dt.out2[type == "meanRisk"], aes(x = time,
group = level))+
  geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
0.3)+
  geom_line(aes(y = estimate, color = level), size = 1)+

  theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
legend.position="top" )+
  scale_x_continuous(breaks=seq(0,500,50)) + scale_y_continuous(labels
= scales::percent)+

  xlab("Days since intiation of treatment")+
  ylab("Absolute Risk of treatment discontinuation (%)")+
  labs(colour="Groups:", fill = "Groups:", title = "Absolute risk of
treatment discontinuation by type of treatment - BARI vs TNFi")+

```

```

1
2
3     labs(group = "Groups:")
4
5
6 plot.ate.diff2
7
8 Second plot to get the ratio in average treatment effect
9
10 plot.ate.ratio2 <- ggplot(dt.out2[type == "ratioRisk"], aes(x = time,
11 group = level))+
12   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
13 0.3)+
14   geom_line(aes(y = estimate, color = level), size = 1)+
15
16   theme_minimal()+
17   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
18   scale_x_continuous(breaks=seq(100,400,50))+
19   scale_y_continuous(limits = c(0.8,3))+
20
21   xlab("Days since initiation of treatment")+
22   ylab("Ratio in Average Treatment Effect")+
23   labs(colour="treatment", fill = "treatment")
24

```

```

25 plot.ate.ratio2
26

```

We can also consider the AIPTW estimate at a specific time point. For example at 365-day.

```

27
28
29 r.one <- dt.out2[type == "diffRisk" & time == 365, .
30 (estimate,lower,upper,p.value)]
31 r.two <- dt.out2[type == "ratioRisk" & time == 365, .
32 (estimate,lower,upper,p.value)]
33
34
35 ploufrows <- c("Difference in average treatment effect","Ratio in
36 average treatment effect")
37 ploufcols <- c("Estimate","95%CI","p")
38 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
39 length(ploufcols))
40 rownames(coxtable) <- ploufrows
41 colnames(coxtable) <- ploufcols
42
43 library(formattable)
44 coxtable[1,"Estimate"] <- paste0(formattable(r.one$estimate*100),"%")
45 coxtable[1,"95%CI"] <-
46 paste0(formattable(r.one$lower),"-",formattable(r.one$upper))
47 coxtable[1,"p"] <- writepvalue(r.one$p.value)
48 coxtable[2,"Estimate"] <- paste0(r.two$estimate)
49 coxtable[2,"95%CI"] <- paste0(r.two$lower,"-",r.two$upper)
50 coxtable[2,"p"] <- writepvalue(r.two$p.value)
51
52

```

```

53 coxtable
54

```

```

55 # Interpretation: If every patient had received BARI, the 365-day risk
56
57
58
59

```

of treatment discontinuation would have been xx% (points) lower compared to when every patient had received TNFi.

.

[3] 1st LINE vs 1st LINE analysis

Common Table 1

Table 1 with NA, to have exact counts and proportions

```

BARI_first <- BARI_DATA
BARI_first <- BARI_first[line_of_therapy == "1st"] # selection of TC
first line

myVars2 <- c("gender", "age_base", "disease_duration_base_years",
"CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
"seropositivity_base", "time_on_drug365", "TC_with_csDMARD",
"line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
"PREDNISON_STEROID_dose", "dose", "initiation_year",
"time_on_drug", "HAQ_score_base")

catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
"obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
"time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
"seropositivity_base")

nonnormalVars <- c()

tab1 <- CreateTableOne(vars = myVars2, data = BARI_first, factorVars =
catVars2, strata = "cohort", test = F, includeNA = T)
tablexp <- print(tab1, nonnormal= nonnormalVars, catDigits = 1,
contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)

saving

write.xlsx(tablexp, file = "./3_clean_output/BARI 3 groups first line
table1 NA.xlsx")

```

Table 1 without NA to have adequate p values to interpret

```

BARI_first <- BARI_DATA
BARI_first <- BARI_first[line_of_therapy == "1st"] # selection TC
first line

myVars2 <- c("gender", "age_base", "disease_duration_base_years",
"CDAI0_raw", "CDAI0", "obese_base", "smoker_base",
"seropositivity_base", "time_on_drug365", "TC_with_csDMARD",

```

```

1
2
3 "line_of_therapy", "N_prev_tsDMARD", "PREDNISON_STEROID",
4 "PREDNISON_STEROID_dose", "dose", "initiation_year",
5 "time_on_drug","HAQ_score_base")
6
7
8 catVars2 <- c("PREDNISON_STEROID", "TC_with_csDMARD", "gender",
9 "obese_base", "smoker_base", "line_of_therapy", "time_on_drugDiff0",
10 "time_on_drug365", "N_prev_tsDMARD", "dose", "initiation_year",
11 "seropositivity_base")
12
13 nonnormalVars <- c()
14
15 tab1 <- CreateTableOne(vars = myVars2, data = BARI_first, factorVars =
16 catVars2, strata = "cohort", test = T, includeNA = F)
17 tablexp <- print(tab1, nonnormal= nonnormalVars, catDigits = 1,
18 contDigits=1, pDigits=2, quote = FALSE, noSpaces = TRUE)
19

```

20 Saving

```

21
22 write.xlsx(tablexp, file = "./3_clean_output/BARI 3 groups first line
23 table1.xlsx")
24
25 summary(BARI_first[cohort=="BARI", c("TC_id", "patient_id",
26 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
27 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
28 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
29 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
30 "cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
31 to see NA values for all variables
32
33
34 summary(BARI_first[cohort=="TNFi", c("TC_id", "patient_id",
35 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
36 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
37 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
38 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
39 "cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
40 to see NA values for all variables
41
42
43 summary(BARI_first[cohort=="OMA", c("TC_id", "patient_id",
44 "stop_DMARD", "stop_reasons", "age_base", "concomitant_csDMARD",
45 "concomitant_csDMARD_type", "TC_with_csDMARD", "PREDNISON_STEROID",
46 "CDAI0", "CDAI0_raw", "disease_duration_base_years", "time_on_drug",
47 "bmi_base", "smoker_base", "line_of_therapy", "obese_base", "gender",
48 "cohort", "adverse_event_reported", "seropositivity_base", "dose")]) #
49 to see NA values for all variables
50

```

51 Non-adjusted Survival curves

52 BARI vs TNFi

```

53
54 BARI_first1 <- copy(BARI_first[cohort %in% c("BARI", "TNFi")])
55
56
57
58
59

```

```

1
2
3 surv_object3 <- Surv(time = BARI_first1$time_on_drug, event =
4 BARI_first1$stop_DMARD) # indiquer stop variable and time_on_drug
5 summary(coxph(surv_object3 ~ cohort, data = BARI_first1))
6 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first1) # fonction
7 which creates Kaplan-meier data
8 survplot_first1 <- ggsurvplot(fit3, data = BARI_first1, # plot
9     pval = T,
10    pval.method = TRUE,
11    legend.title = "Groups :",
12    legend.labs = c("Baricitinib", "TFNi"),
13    xlab = "Time (days)",
14    xlim = c(0, 700),
15    censor = FALSE,
16    title = "Non-adjusted model of drug discontinuation by type
17 of treatment",
18    surv.median.line = "v",
19    linetype = 1,
20    size = 1.5,
21    ggtheme = theme_minimal(),
22    # palette = c("grey78", "grey50", "grey10"),
23    palette = c("red2", "green3"), # to get colors
24    risk.table = T
25    )
26
27
28
29 survplot_first1
30 table(BARI_first1$cohort)
31 summary(fit3)
32
33 rm(surv_object3, fit3)
34
35 BARI vs OMA
36
37 BARI_first2 <- BARI_first[line_of_therapy == "1st" & cohort %in%
38 c("BARI", "OMA")] # selection des TC TNFi
39
40 surv_object3 <- Surv(time = BARI_first2$time_on_drug, event =
41 BARI_first2$stop_DMARD) # indiquer stop variable and time_on_drug
42 summary(coxph(surv_object3 ~ cohort, data = BARI_first2))
43 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first2) # fonction
44 which creates Kaplan-meier data
45 survplot_first2 <- ggsurvplot(fit3, data = BARI_first2, # plot
46     pval = T,
47    pval.method = TRUE,
48    legend.title = "Groups :",
49    legend.labs = c("Baricitinib", "OMA"),
50    xlab = "Time (days)",
51    xlim = c(0, 700),
52    censor = FALSE,
53    title = "Non-adjusted model of drug discontinuation by type
54 of treatment",
55
56
57
58
59

```

```

1
2
3     surv.median.line = "v",
4     linetype = 1,
5     size = 1.5,
6     ggtheme = theme_minimal(),
7     # palette = c("grey78", "grey50", "grey10"),
8     palette = c("red2", "blue3"), # to get colors
9     risk.table = T
10    )
11

```

```

12
13 survplot_first2
14 table(BARI_first2$cohort)
15 summary(fit3)
16

```

```

17 rm(surv_object3, fit3)
18

```

Adjusted survival analyses

BARI vs TNFi

Verification (quick)

```

25 # Test of proportionality of hazards on raw data
26 test_first_ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
27 as.factor(cohort)+
28                          cluster(patient_id),
29                          data= BARI_first1)
30
31 cox.zph(test_first_ph)
32

```

Adjusted Cox-model

```

33 imputed_data1_first <- complete(imputed_data1,"long",include=T) # to
34 put in the long format
35 imputed_data1_first <- filter(imputed_data1_first, line_of_therapy ==
36 "1st") # only keep 1st line imputed TC
37 imputed_data1_first <- as.mids(imputed_data1_first) # put back in
38 previous format, to use fit.mult.impute
39
40
41 BARI_first1.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
42 cohort+

```

```

43                          I(age_base/10)+
44                          bmi_base+
45                          concomitant_csDMARD+
46                          PREDNISON_STEROID+
47                          I(CDAI0/10)+
48                          I(disease_duration_base_years/10)+
49                          C(smoker_base, base=3)+
50                          line_of_therapy+
51                          gender+
52                          seropositivity_base+
53                          cluster(patient_id),
54                          fitter = coxph, xtrans =
55
56
57
58
59
60

```

```

1
2
3 imputed_data1_first, data = BARI_first1)
4 summary(BARI_first1.adj.mi)
5
6 Creation of HR table with p-values
7 ploufrows <- names(BARI_first1.adj.mi$coefficients)
8 ploufcols <- c("HR","95%CI","p")
9 coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
10 length(ploufcols))
11 rownames(coxtable) <- ploufrows
12 colnames(coxtable) <- ploufcols
13 plouf <- summary(BARI_first1.adj.mi)
14
15
16 for(row in ploufrows)
17 {
18   coxtable[row,"HR"] <-
19   formattable(plouf$coefficients[row,"exp(coef)"])
20   coxtable[row,"95%CI"] <-
21   paste0(formattable(plouf$conf.int[row,"lower .95"]),"-",formattable(pl
22 ouf$conf.int[row,"upper .95"]))
23   coxtable[row,"p"] <- writepvalue(plouf$coefficients[row,"Pr(>|
24 z|)"])
25 }
26
27
28 write.xlsx(coxtable, file="./3_clean_output/BARI vs TNFi HR 1st
29 lines.xlsx") # save in excel format
30
31 Adjusted curves with imputed data
32 dummy_cox_impute_first1 <- mice::complete(imputed_data1_first, "long",
33 include = T)
34 dummy_cox_impute_first1 <-
35 dummy_cox_impute_first1[dummy_cox_impute_first1$.imp != 0,]
36
37 BARI_first1_fit <- survfit(coxph(Surv(time = time_on_drug, event =
38 stop_DMARD) ~ cohort+
39
40 I(age_base/10)+
41 bmi_base+
42 concomitant_csDMARD+
43 PREDNISON_STEROID+
44 CDAI0+
45 I(disease_duration_base_years/10)+
46 C(smoker_base, base=3)+
47 line_of_therapy+
48 gender+
49 seropositivity_base+
50 cluster(patient_id)+
51 strata(cohort),dummy_cox_impute_first1),
52 data = dummy_cox_impute_first1)
53
54
55
56 survplot_first1_adj <- ggsurvplot(BARI_first1_fit, data =
57
58
59
60

```



```

1
2
3 dummy_cox_impute_first1, variable = "cohort",
4     xlab = "Time (days)",
5     title = "Multivariable Cox model of drug discontinuation by
6 type of treatment - 1st line vs 1st line",
7     legend.title = "Groups :",
8     legend.labs = c("Baricitinib", "TNFi"),
9     censor = FALSE,
10    xlim = c(0, 700),
11    surv.median.line = "v",
12    linetype = 1,
13    size = 1.5,
14    ggtheme = theme_minimal(),
15    #palette = c("grey78", "grey10")
16    palette = c("red2", "green3"), # to get colors
17
18 )
19

```

```

20 survplot_first1_adj <- survplot_first1_adj +
21     labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
22 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
23 smoking status, line of therapy, gender, seropositivity")
24

```

```

25
26 survplot_first1_adj
27 table(BARI_first1$cohort)
28 rm(dummy_cox_impute_first1, BARI_first1_fit)
29

```

BARI vs OMA

Verification (quick)

```

31
32
33 # Test of proportionality of hazards on raw data
34 test_first_ph <- coxph(Surv(time = time_on_drug, event = stop_DMARD) ~
35 as.factor(cohort)+
36
37     cluster(patient_id),
38     data= BARI_first2)
39 cox.zph(test_first_ph)
40

```

Adjusted Cox-model

```

41
42 imputed_data2_first <- complete(imputed_data2,"long",include=T) # to
43 put in the long format
44 imputed_data2_first <- filter(imputed_data2_first, line_of_therapy ==
45 "1st") # only keep 1st line imputed TC
46 imputed_data2_first <- as.mids(imputed_data2_first) # put back in
47 previous format, to use fit.mult.impute
48

```

```

49
50 BARI_first2.adj.mi <- fit.mult.impute(Surv(time_on_drug, stop_DMARD) ~
51 cohort+
52     I(age_base/10)+
53     bmi_base+
54     concomitant_csDMARD+
55     PREDNISON_STEROID+
56
57
58
59

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

I(CDAI0/10)+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+
gender+
seropositivity_base+
cluster(patient_id),
fitter = coxph, xtrans =

```

imputed_data2_first, data = BARI_first2)
summary(BARI_first2.adj.mi)

```

Creation of HR table with p-values

```

ploufrows <- names(BARI_first2.adj.mi$coefficients)
ploufcols <- c("HR", "95%CI", "p")
coxtable <- matrix(data = NA, nrow = length(ploufrows), ncol =
length(ploufcols))
rownames(coxtable) <- ploufrows
colnames(coxtable) <- ploufcols
plouf <- summary(BARI_first2.adj.mi)

```

```

for(row in ploufrows)
{
  coxtable[row, "HR"] <-
formattable(plouf$coefficients[row, "exp(coef)"])
  coxtable[row, "95%CI"] <-
paste0(formattable(plouf$conf.int[row, "lower .95"]), "-", formattable(pl
ouf$conf.int[row, "upper .95"]))
  coxtable[row, "p"] <- writepvalue(plouf$coefficients[row, "Pr(>|
z|)"])
}

```

```

write.xlsx(coxtable, file="./3_clean_output/BARI vs OMA HR 1st
lines.xlsx") # save in excel format

```

Adjusted curves with imputed data

```

dummy_cox_impute_first2 <- mice::complete(imputed_data2_first, "long",
include = T)
dummy_cox_impute_first2 <-
dummy_cox_impute_first2[dummy_cox_impute_first2$.imp != 0,]

```

```

BARI_first2_fit <- survfit(coxph(Surv(time = time_on_drug, event =
stop_DMARD) ~ cohort+

```

I(age_base/10)+
bmi_base+
concomitant_csDMARD+
PREDNISON_STEROID+
CDAI0+
I(disease_duration_base_years/10)+
C(smoker_base, base=3)+
line_of_therapy+

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

      gender+
      seropositivity_base+
      cluster(patient_id)+
      strata(cohort), dummy_cox_impute_first2),
data = dummy_cox_impute_first2)

survplot_first2_adj <- ggsurvplot(BARI_first2_fit, data =
dummy_cox_impute_first2, variable = "cohort",
  xlab = "Time (days)",
  title = "Multivariable Cox model of drug discontinuation by
type of treatment - 1st line vs 1st line",
  legend.title = "Groups :",
  legend.labs = c("Baricitinib", "OMA bDMARDs"),
  censor = FALSE,
  xlim = c(0, 700),
  surv.median.line = "v",
  linetype = 1,
  size = 1.5,
  ggtheme = theme_minimal(),
  #palette = c("grey78", "grey50")
  palette = c("red2", "blue3"), # to get colors
)

survplot_first2_adj <- survplot_first2_adj +
  labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
concomitant glucocorticoid, baseline CDAI, disease duration (decades),
smoking status, line of therapy, gender, seropositivity")

survplot_first2_adj
table(BARI_first2$cohort)
rm(dummy_cox_impute_first2, BARI_first2_fit)

```

1. [4] LACK of EFFICACY and ADVERSE EVENTS

Analysis by stop_reasons in competing risk

(BARI vs TNFi)

Cumulative incidence function

```
BARI_comp <- copy(BARI_DATA)
```

#General

```

BARI_comp[stop_reasons == "ADVERSE_EVENT", status := 1]
BARI_comp[stop_reasons == "NOT_EFFECTIVE", status := 2]
BARI_comp[stop_reasons == "OTHER" | stop_reasons == "REMISSION",
status := 3]

```

```

1
2
3 BARI_comp[stop_reasons == "CONTINUE", status := 0]
4 BARI_comp$cohort <- as.factor(BARI_comp$cohort)
5
6 library(reshape)
7
8 BARI_comp_B <- BARI_comp[cohort %in% c("BARI")] #BARI only
9 ci_BARI <- Cuminc(time = "time_on_drug", status = "status", data =
10 BARI_comp_B)
11 ci_BARI <- ci_BARI[, -c(2,6,7,8,9)]
12 ci_long_BARI <- reshape2::melt(ci_BARI, id.vars = "time")
13
14 BARI_comp_T <- BARI_comp[cohort %in% c("TNFi")] #TNFi only
15 ci_TNFi <- Cuminc(time = "time_on_drug", status = "status", data =
16 BARI_comp_T)
17 ci_TNFi <- ci_TNFi[, -c(2,6,7,8,9)]
18 ci_long_TNFi <- reshape2::melt(ci_TNFi, id.vars = "time")
19
20
21 ci_long_BARI$cohort <- 0
22 ci_long_TNFi$cohort <- 1
23 ci_long <- rbind(ci_long_BARI, ci_long_TNFi)
24 ci_long$cohort <- as.factor(ci_long$cohort)
25
26
27 plot2 <- ggplot(data = ci_long, aes(x = time,
28 y = value,
29 linetype =
30 interaction(cohort, variable),
31 col =
32 interaction(cohort, variable))) +
33 geom_line(size = 0.75) +
34 scale_color_manual(name = "",
35 values
36 =c("#08306B", "#08306B", "#238B45", "#238B45", "#FD8D3C", "#FD8D3C"),
37 labels = c("Adverse Event (BARI)", "Adverse Event
38 (TNFi)", "Ineffectiveness (BARI)", "Ineffectiveness (TNFi)", "Other
39 (BARI)", "Other (TNFi)"))+
40 scale_linetype_manual(name="",
41 values = c(1,3,1,3,1,3),
42 labels = c("Adverse Event (BARI)", "Adverse
43 Event (TNFi)", "Ineffectiveness (BARI)", "Ineffectiveness (TNFi)", "Other
44 (BARI)", "Other (TNFi)"))+
45 scale_x_continuous(name = "Time", limits = c(1,365)) +
46 scale_y_continuous(name = "Cumulative incidence", limits =
47 c(0.0,0.3)) +
48 theme_bw()+
49 theme(strip.text.y = element_blank(),
50 strip.background = element_blank(),
51 axis.line.x = element_line(size = 0.5),
52 axis.text = element_text(face = "bold", colour = "black"),
53 legend.position="right", plot.margin =
54 unit(c(1,3,2,1),"lines"))+
55
56
57
58
59
60

```

```
1
2
3     #ggtitle("Cumulative incidence functions")+
4     theme(plot.title = element_text(hjust = 0.5))
5
6 plot2
7
8 Adjusting variables
9 # Covariates of interest for Cox
10
11 covs <-
12 c("cohort","age_base","bmi_base","TC_with_csDMARD","PREDNISON_STEROID"
13   ,"CDAI0","disease_duration_base_years","smoker_base","line_of_therapy"
14   ,"gender","seropositivity_base")
15
16 Cause-specific hazard model
17 # Rappel: imputed_data1 = BARI vs TNFi
18 #         imputed_data2 = BARI vs OMA
19
20
21 # Transition matrix definition
22 tmat <- trans.comprisk(2, names = c("event-free","ae","lae"))
23 tmat
24
25
26 imputed_data1_long <- complete(imputed_data1, action = "long") %>%
27 setDT()
28 imputed_data1_long[,stop_ae := fifelse(stop_reasons ==
29   "ADVERSE_EVENT",1,0)]
30 imputed_data1_long[,stop_lae := fifelse(stop_reasons ==
31   "NOT_EFFECTIVE",1,0)]
32 imputed_data1_long[,stop_other := fifelse(stop_reasons == "OTHER" |
33   stop_reasons == "REMISSION",1,0)]
34 #[,continue := fifelse(stop_reasons == "CONTINUE",1,0)]
35 imputed_data1_long[,continue := fifelse(stop_reasons == "OTHER" |
36   stop_reasons == "REMISSION" | stop_reasons == "CONTINUE",1,0)]
37
38 M <- imputed_data1$m
39
40
41 mice_fit <- lapply(1:M,function(m){
42
43   # subset
44   data_sub <- imputed_data1_long[.imp == m]
45
46   mst_hosp <- msprep(time =
47 c("time_on_drug","time_on_drug","time_on_drug"),
48   status = c("continue","stop_ae","stop_lae"),
49   data = as.data.frame(data_sub),
50   trans = tmat,
51   keep = covs)
52
53   # get covariates
54 tmp <- expand.covs(mst_hosp,covs, append = TRUE, longnames = T)
55 tmp_cov <- grep(paste0(covs,".",collapse = "|"),names(tmp),value = T)
56
57
58
59
60
```

```

1
2
3   # fit
4   coxph(as.formula(paste0("Surv(Tstart, Tstop, status) ~",
5                           paste0(tmp_cov,collapse = " + "),
6                           "+ strata(trans)")),
7
8       data = tmp,
9       method = "breslow")
10
11  }) %>%
12    as.mira()
13
14  est <- pool(mice_fit)
15
16  # Transition 1 = Adverse Event. Hazard Ratio vs VARI
17  # Transition 2 = Lack of Efficacy. Hazard Ratio vs VARI
18  # estimate = Hazard ratio
19
20  summary(est, conf.int = T, exponentiate = T)
21
22
23  # Conclusion
24  # => The hazard ratio of lack of efficacy (lae) for TNFi is 65% higher
25  than for BARI. Significant.
26  # => No difference between TNFi and BARI for Adverse Event (ae)
27
28  Clean table with confidence intervals & p-values
29
30  # Hazard ratios
31  ploufrows <- as.character(summary(est)$term)
32
33  ploufcols <- c("HR","95%CI","p")
34  coxtable_csh <- matrix(data = NA, nrow = length(ploufrows), ncol =
35  length(ploufcols))
36  rownames(coxtable_csh) <- ploufrows
37  colnames(coxtable_csh) <- ploufcols
38  plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
39
40  for(row in ploufrows)
41  {
42    coxtable_csh[row,"HR"] <- formattable(plouf[term %in% row,
43  estimate])
44    coxtable_csh[row,"95%CI"] <- paste0(formattable(plouf[term %in%
45  row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
46    coxtable_csh[row,"p"] <- writepvalue(plouf[term %in% row,
47  p.value])}
48
49
50  output <- coxtable_csh
51  row.names(output)[1:2] <- c("TNFi Adverse Event (vs BARI)", "TNFi Lack
52  of Eff (vs BARI)")
53
54
55  # Transition 1 = Adverse Event. Hazard Ratio vs VARI
56
57
58
59
60

```

```
1
2
3 # Transition 2 = Lack of Efficacy. Hazard Ratio vs VARI
4 output
5
6 Subdistribution hazard model (Fine-Gray)
7 # Status variable
8 imputed_data1_long[stop_reasons == "ADVERSE_EVENT",status := 1]
9 imputed_data1_long[stop_reasons == "NOT_EFFECTIVE", status := 2]
10 imputed_data1_long[stop_reasons == "OTHER" | stop_reasons ==
11 "REMISSION", status := 3]
12 imputed_data1_long[stop_reasons == "CONTINUE", status := 0]
13
14
15 ## ATTENTION levels() re-ecrit juste l'étiquette!! Change pas la
16 donnée !!! Donc ça re écrit les labels
17
18 imputed_data1_long$line_of_therapy <-
19 as.factor(imputed_data1_long$line_of_therapy)
20 imputed_data1_long$seropositivity_base <-
21 as.factor(imputed_data1_long$seropositivity_base)
22
23
24 levels(imputed_data1_long$cohort) <- c("0","1")
25 levels(imputed_data1_long$line_of_therapy) <- c("0","1","2","3")
26 levels(imputed_data1_long$gender) <- c("0","1")
27 levels(imputed_data1_long$smoker_base) <- c("2","1","0")
28 levels(imputed_data1_long$smoker_base)
29 levels(imputed_data1_long$seropositivity_base) <- c("0","1")
30
31 M <- imputed_data1$m
32
33 # First loop to get estimates for event = 1: ADVERSE EVENT
34
35 mice_fit <- lapply(1:M,function(m){
36   # subset
37   BARI_toto <- imputed_data1_long[.imp == m]
38   # subdistribution hazard model
39   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
40 BARI_toto[,..covs],failcode = 1,cencode = 0)
41
42
43 }) %>%
44   as.mira()
45 est <- pool(mice_fit)
46 summary(est, conf.int = T, exponentiate = T)
47
48 # Second loop to get estimates for event = 2: LACK OF EFFICACY
49
50 mice_fit2 <- lapply(1:M,function(m){
51   # subset
52   BARI_toto <- imputed_data1_long[.imp == m]
53   #subdistribution hazard model
54   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
```

```

1
2
3   BARI_toto[,..covs],failcode = 2,cencode = 0)
4
5   }) %>%
6     as.mira()
7
8
9   est2 <- pool(mice_fit2)
10  summary(est2, conf.int = T, exponentiate = T)
11
12  # Conclusions:
13  # No significant difference in incidence of adverse event between TNFi
14  and BARI
15  # Increased incidence of lack of efficacy for TNFi compared to BARI.
16  # CAREFUL: using a Fine-Gray model allows us to make claim about the
17  association between a covariate and the direction of the increase in
18  incidence, but we can't quantify the magnitude of the increase in
19  incidence.
20
21  Clean tables with Hazard ratios with confidence intervals & p-values
22
23  # Adverse event
24  ploufrows <- as.character(summary(est)$term)
25  ploufcols <- c("HR","95%CI","p")
26  coxtable_ae <- matrix(data = NA, nrow = length(ploufrows), ncol =
27  length(ploufcols))
28  rownames(coxtable_ae) <- ploufrows
29  colnames(coxtable_ae) <- ploufcols
30  plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
31
32
33  for(row in ploufrows)
34  {
35    coxtable_ae[row,"HR"] <- formattable(plouf[term %in% row,
36  estimate])
37    coxtable_ae[row,"95%CI"] <- paste0(formattable(plouf[term %in% row,
38  `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
39    coxtable_ae[row,"p"] <- writepvalue(plouf[term %in% row, p.value])}
40
41  row.names(coxtable_ae)[1] <- c("TNFi vs BARI Advserse Events")
42
43
44  # Lack of efficacy
45  ploufrows <- as.character(summary(est2)$term)
46  ploufcols <- c("HR","95%CI","p")
47  coxtable_lae <- matrix(data = NA, nrow = length(ploufrows), ncol =
48  length(ploufcols))
49  rownames(coxtable_lae) <- ploufrows
50  colnames(coxtable_lae) <- ploufcols
51  plouf <- summary(est2, conf.int = T, exponentiate = T) %>% setDT()
52
53
54  for(row in ploufrows)
55  {
56    coxtable_lae[row,"HR"] <- formattable(plouf[term %in% row,
57

```



```

1
2
3 estimate])
4   coxtable_lae[row,"95%CI"] <- paste0(formatable(plouf[term %in%
5 row, `2.5 %`]),"-",formatable(plouf[term %in% row, `97.5 %`]))
6   coxtable_lae[row,"p"] <- writepvalue(plouf[term %in% row,
7 p.value])}
8
9
10 row.names(coxtable_lae)[1] <- c("TNFi vs BARI Lack of Eff")
11
12 # output
13 coxtable_ae
14 coxtable_lae
15
16 write.xlsx(coxtable_ae, file="./3_clean_output/BARI vs TNFi HR
17 competing risk Fine-Gray AE.xlsx") # saving excel file
18 write.xlsx(coxtable_lae, file="./3_clean_output/BARI vs TNFi HR
19 competing risk Fine-Gray LAE.xlsx") # saving excel file
20
21 (BARI vs OMA)
22
23 Cumulative incidence function
24 BARI_comp <- copy(BARI_DATA)
25
26 #General
27
28
29 BARI_comp[stop_reasons == "ADVERSE_EVENT",status := 1]
30 BARI_comp[stop_reasons == "NOT_EFFECTIVE", status := 2]
31 BARI_comp[stop_reasons == "OTHER" | stop_reasons == "REMISSION",
32 status := 3]
33 BARI_comp[stop_reasons == "CONTINUE", status := 0]
34 BARI_comp$cohort <- as.factor(BARI_comp$cohort)
35
36
37 library(reshape)
38
39 BARI_comp_B <- BARI_comp[cohort %in% c("BARI")] #BARI only
40 ci_BARI <- Cuminc(time = "time_on_drug",status = "status", data =
41 BARI_comp_B)
42 ci_BARI <- ci_BARI[, -c(2,6,7,8,9)]
43 ci_long_BARI <- reshape2::melt(ci_BARI,id.vars = "time")
44
45
46 BARI_comp_O <- BARI_comp[cohort %in% c("OMA")] #OMA only
47 ci_OMA <- Cuminc(time = "time_on_drug",status = "status", data =
48 BARI_comp_O)
49 ci_OMA <- ci_OMA[, -c(2,6,7,8,9)]
50 ci_long_OMA <- reshape2::melt(ci_OMA,id.vars = "time")
51
52 ci_long_BARI$cohort <- 0
53 ci_long_OMA$cohort <- 1
54 ci_long_2 <- rbind(ci_long_BARI,ci_long_OMA)
55 ci_long_2$cohort <- as.factor(ci_long_2$cohort)
56
57
58
59
60

```

```

1
2
3
4 plot3 <- ggplot(data = ci_long_2, aes(x = time,
5                                     y = value,
6                                     linetype =
7                                     interaction(cohort,variable),
8                                     col =
9                                     interaction(cohort,variable))) +
10   geom_line(size = 0.75) +
11   scale_color_manual(name = "",
12                     values =
13   c("#08306B", "#08306B", "#238B45", "#238B45", "#FD8D3C", "#FD8D3C"),
14   labels = c("Adverse Event (BARI)", "Adverse Event
15 (OMA)", "Ineffectiveness (BARI)", "Ineffectiveness (OMA)", "Other
16 (BARI)", "Other (OMA)"))+
17   scale_linetype_manual(name="",
18   values = c(1,3,1,3,1,3),
19   labels = c("Adverse Event (BARI)", "Adverse
20 Event (OMA)", "Ineffectiveness (BARI)", "Ineffectiveness (OMA)", "Other
21 (BARI)", "Other (OMA)"))+
22   scale_x_continuous(name = "Time", limits = c(1,365)) +
23   scale_y_continuous(name = "Cumulative incidence", limits =
24 c(0.0,0.3)) +
25   theme_bw()+
26   theme(strip.text.y = element_blank(),
27         strip.background = element_blank(),
28         axis.line.x = element_line(size = 0.5),
29         axis.text = element_text(face = "bold", colour = "black"),
30         legend.position="right", plot.margin =
31   unit(c(1,3,2,1),"lines"))+
32   #ggtitle("Cumulative incidence functions")+
33   theme(plot.title = element_text(hjust = 0.5))
34
35
36

```

```

37 plot3
38

```

Adjusting variables

```

39 # Covariates of interest for Cox
40

```

```

41
42 covs <-
43 c("cohort", "age_base", "bmi_base", "TC_with_csDMARD", "PREDNISON_STEROID"
44 , "CDAI0", "disease_duration_base_years", "smoker_base", "line_of_therapy"
45 , "gender", "seropositivity_base")
46

```

Cause-specific hazard model

```

47
48 # Rappel: imputed_data1 = BARI vs TNFi
49 #           imputed_data2 = BARI vs OMA
50

```

```

51 # Transition matrix definition
52

```

```

53 library(mstate)
54 tmat <- trans.comprisk(2, names = c("event-free", "ae", "lae"))
55 tmat
56
57
58
59

```

```
1
2
3
4 imputed_data2_long <- complete(imputed_data2, action = "long") %>%
5 setDT()
6 imputed_data2_long[,stop_ae := fifelse(stop_reasons ==
7 "ADVERSE_EVENT",1,0)]
8 imputed_data2_long[,stop_lae := fifelse(stop_reasons ==
9 "NOT_EFFECTIVE",1,0)]
10
11 imputed_data2_long[,stop_other := fifelse(stop_reasons == "OTHER" |
12 stop_reasons == "REMISSION",1,0)]
13 #[,continue := fifelse(stop_reasons == "CONTINUE",1,0)]
14 imputed_data2_long[,continue := fifelse(stop_reasons == "OTHER" |
15 stop_reasons == "REMISSION" | stop_reasons == "CONTINUE",1,0)]
16
17 M <- imputed_data2$m
18
19 mice_fit <- lapply(1:M,function(m){
20
21   # subset
22   data_sub <- imputed_data2_long[.imp == m]
23
24   mst_hosp <- msprep(time =
25 c("time_on_drug","time_on_drug","time_on_drug"),
26                       status = c("continue","stop_ae","stop_lae"),
27                       data = as.data.frame(data_sub),
28                       trans = tmat,
29                       keep = covs)
30
31   # get covariates
32 tmp <- expand.covs(mst_hosp,covs, append = TRUE, longnames = T)
33 tmp_cov <- grep(paste0(covs,".",collapse = "|"),names(tmp),value = T)
34
35   # fit
36 coxph(as.formula(paste0("Surv(Tstart, Tstop, status) ~",
37                         paste0(tmp_cov,collapse = " + "),
38                         "+ strata(trans)")),
39       data = tmp,
40       method = "breslow")
41
42 }) %>%
43   as.mira()
44
45
46 est <- pool(mice_fit)
47 summary(est, conf.int = T, exponentiate = T)
48
49 # Transition 1 = Adverse Event
50 # Transition 2 = Lack of Efficacy
51
52
53 # Conclusion
54 # => No difference between OMA and BARI for Adverse Event (ae) and for
55 Lack of Event (lae)
56
57
58
59
60
```

Cleaner table with Hazard ratios with confidence intervals & p-values

```

1
2
3
4
5 ploufrows <- as.character(summary(est)$term)
6 ploufcols <- c("HR", "95%CI", "p")
7 coxtable_csh2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
8 length(ploufcols))
9 rownames(coxtable_csh2) <- ploufrows
10 colnames(coxtable_csh2) <- ploufcols
11 plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
12
13
14 for(row in ploufrows)
15 {
16   coxtable_csh2[row, "HR"] <- formattable(plouf[term %in% row,
17 estimate])
18   coxtable_csh2[row, "95%CI"] <- paste0(formattable(plouf[term %in%
19 row, `2.5 %`]), "-", formattable(plouf[term %in% row, `97.5 %`]))
20   coxtable_csh2[row, "p"] <- writepvalue(plouf[term %in% row,
21 p.value])}
22
23 row.names(coxtable_csh2)[1:2] <- c("OMA vs BARI Adverse event", "OMA
24 vs BARI Lack of Eff" )
25 coxtable_csh2
26

```

Subdistribution hazard model (Fine-Gray)

```

27
28 # Status variable
29 imputed_data2_long[stop_reasons == "ADVERSE_EVENT", status := 1]
30 imputed_data2_long[stop_reasons == "NOT_EFFECTIVE", status := 2]
31 imputed_data2_long[stop_reasons == "OTHER" | stop_reasons ==
32 "REMISSION", status := 3]
33 imputed_data2_long[stop_reasons == "CONTINUE", status := 0]
34
35
36 imputed_data2_long$line_of_therapy <-
37 as.factor(imputed_data2_long$line_of_therapy)
38 imputed_data2_long$seropositivity_base <-
39 as.factor(imputed_data2_long$seropositivity_base)
40 levels(imputed_data2_long$cohort) <- c("0", "1")
41 levels(imputed_data2_long$line_of_therapy) <- c("0", "1", "2", "3")
42 levels(imputed_data2_long$gender) <- c("0", "1")
43 levels(imputed_data2_long$smoker_base) <- c("2", "1", "0")
44 levels(imputed_data2_long$smoker_base)
45 levels(imputed_data2_long$seropositivity_base) <- c("0", "1")
46
47

```

```

48 M <- imputed_data2$m
49

```

```

50 # First loop to get estimates for event = 1: ADVERSE EVENT
51

```

```

52 mice_fit <- lapply(1:M, function(m){
53   # subset
54   BARI_toto <- imputed_data2_long[.imp == m]
55
56
57
58
59

```

```

1
2
3     #subdistribution hazard model
4     shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
5 BARI_toto[,..covs],failcode = 1,cencode = 0)
6
7   }) %>%
8     as.mira()
9   est <- pool(mice_fit)
10  summary(est, conf.int = T, exponentiate = T)
11
12
13 # Second loop to get estimates for event = 2: LACK OF EFFICACY
14 mice_fit2 <- lapply(1:M,function(m){
15
16   # subset
17   BARI_toto <- imputed_data2_long[.imp == m]
18
19   # subdistribution hazard model
20   shm <- crr(BARI_toto$time_on_drug,BARI_toto$status,cov1 =
21 BARI_toto[,..covs],failcode = 2,cencode = 0)
22
23 }) %>%
24   as.mira()
25   est2 <- pool(mice_fit2)
26   summary(est2, conf.int = T, exponentiate = T)
27
28
29 # Conclusions:
30 # No significant difference in incidence of "adverse event" and "lack
31 of efficacy" between TNFi and BARI
32
33 # CAREFUL: using a Fine-Gray model allows us to make claim about the
34 association between a covariate and the direction of the increase in
35 incidence, but we can't quantify the magnitude of the increase in
36 incidence.
37
38
39 Cleaner Table with Hazard ratios with confidence intervals & p-values
40
41 # Adverse event
42 ploufrows <- as.character(summary(est)$term)
43 ploufcols <- c("HR","95%CI","p")
44 coxtable_ae2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
45 length(ploufcols))
46 rownames(coxtable_ae2) <- ploufrows
47 colnames(coxtable_ae2) <- ploufcols
48 plouf <- summary(est, conf.int = T, exponentiate = T) %>% setDT()
49
50 for(row in ploufrows)
51 {
52   coxtable_ae2[row,"HR"] <- formattable(plouf[term %in% row,
53 estimate])
54   coxtable_ae2[row,"95%CI"] <- paste0(formattable(plouf[term %in%
55 row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
56
57
58
59
60

```

```
1
2
3     coxtable_ae2[row,"p"] <- writepvalue(plouf[term %in% row,
4 p.value])}
5
6     row.names(coxtable_ae2)[1] <- c("OMA vs BARI Advserere Events")
7
8
9     # Lack of efficacy
10    ploufrows <- as.character(summary(est2)$term)
11    ploufcols <- c("HR","95%CI","p")
12    coxtable_lae2 <- matrix(data = NA, nrow = length(ploufrows), ncol =
13 length(ploufcols))
14    rownames(coxtable_lae2) <- ploufrows
15    colnames(coxtable_lae2) <- ploufcols
16    plouf <- summary(est2, conf.int = T, exponentiate = T) %>% setDT()
17
18    for(row in ploufrows)
19    {
20      coxtable_lae2[row,"HR"] <- formattable(plouf[term %in% row,
21 estimate])
22      coxtable_lae2[row,"95%CI"] <- paste0(formattable(plouf[term %in%
23 row, `2.5 %`]),"-",formattable(plouf[term %in% row, `97.5 %`]))
24      coxtable_lae2[row,"p"] <- writepvalue(plouf[term %in% row,
25 p.value])}
26
27
28    row.names(coxtable_lae2)[1] <- c("OMA vs BARI Lack of Eff")
29
30    #Output
31    coxtable_ae2
32    coxtable_lae2
33
34
35    write.xlsx(coxtable_ae2, file="./3_clean_output/BARI vs OMA HR
36 competing risk Fine-Gray AE.xlsx") # saving excel file
37    write.xlsx(coxtable_lae2, file="./3_clean_output/BARI vs OMA HR
38 competing risk Fine-Gray LAE.xlsx") # saving excel file
39
40
```

1. Saving

```
41
42 save.image(file="./3_clean_output/full_workspaces/workspace_1.RData")
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
```

2 - LDA and REM ANALYSIS

10/11/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)

rm(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext())$path))

load("../1_datamanaged_files/datamanaged.Rdata")
```

This code aims at providing estimates for the remission rates of the different treatments groups REM = REMission LDA = Low Disease Activity

Both outcome are base on the CDAI CDAI = Clinical Disease Activity Index

CDAI is an index computed by the physician, which scores the severity of the disease.

1. [0] Exploration

See all available raw CDAI measures :

```
BARI_long[, group := "non-BARI"]
BARI_long[drug == "BIOLOGIC_BARICITINIB", group := "BARI"]

nrow(BARI_DATA)
summary(BARI_DATA[, .(CDAI0_raw, CDAI12_raw)])
```

1. [1] CARRAC (confirm covariates for confounding and for attrition)

For LDA with updated function

```
library(modules)
source_comp_eff <- modules::use("ETAPE_2_supp_code.R")

LDA_BARI_TNF <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "TNFi")],
  var = "CDAI12",
  thres = 10,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons" ),
  seed = 123)

LDA_BARI_OMA <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "OMA")],
  var = "CDAI12",
  thres = 10,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons"),
  seed = 123)

LDA_BARI_TNF
LDA_BARI_OMA
```

For REM with updated function

```
REM_BARI_TNF <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "TNFi")],
  var = "CDAI12",
  thres = 2.8,
  ttt_var = "cohort",
  ref_ttt = "BARI",
  counfunders = c("TC_with_csDMARD", "PREDNISON_STEROID",
                  "line_of_therapy", "CDAI0"),
  attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
                "line_of_therapy", "CDAI0", "stop_reasons" ),
  seed = 123)

REM_BARI_OMA <- source_comp_eff$CARRAC(
  datain = BARI_DATA[cohort %in% c("BARI", "OMA")],
```



```
1
2
3   var = "CDAI12",
4   thres = 2.8,
5   ttt_var = "cohort",
6   ref_ttt = "BARI",
7   counfounders = c("TC_with_csDMARD", "PREDNISON_STEROID",
8                   "line_of_therapy", "CDAI0"),
9   attrition = c("TC_with_csDMARD", "PREDNISON_STEROID",
10              "line_of_therapy", "CDAI0", "stop_reasons"),
11   seed = 123)
```

```
13
14 REM_BARI_TNF
15 REM_BARI_OMA
```

16
17 This methods was developed by Mongin et al,
18 <https://ard.bmj.com/content/early/2022/01/12/annrheumdis-2021-221477>

20 Pooled table

```
21 table <- rbind(LDA_BARI_TNF, LDA_BARI_OMA, REM_BARI_TNF, REM_BARI_OMA)
22
23 write.xlsx(table, file = "./3_clean_output/table_LDA_REM_CARRAC.xlsx",
24            row.names = F)
```

27 1. Saving

```
28
29 save.image(file="./3_clean_output/full_workspaces/workspace_2.RData")
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

2 - LDA and REM supp CODE

10/11/2020

```

1  {r setup, include=FALSE} import("data.table") import("plyr")
2
3
4
5
6
7
8
9
10 {r setup, include=FALSE} import("data.table") import("plyr")
11 import("data.table") import("mice") import("ipw") import("survey")
12 import("geepack") import("futile.logger") import("emmeans")
13 import("stats") import("survival")
14
15

```

function to perform checks on data

```

16
17
18 ``{r setup, include=FALSE} check_data = function(datain, var = "CDAI_fu", ttt_var = "ttt",
19 ref_ttt = "ttt_ref", ID_ttt = NULL, othervar = c())
20
21 {
22
23 data <- setDT(copy(datain))
24
25 vartochek <- Reduce(union,list(var,ttt_var,othervar)) notindata <-
26 setdiff(vartochek,names(data))
27
28 if(length(notindata)>0){ stop(paste0("the variables",paste0(notindata,collapse = ",")," are
29 not in the dataplease correct")) }
30
31 # force ttt as var name setnames(data,ttt_var,"ttt")
32
33 if( data[,uniqueN(ttt)]>2){ stop("there are more than two treatments. The analysis has
34 been implemented only for 2 treatments") }
35
36 if(!any(data$ttt == ref_ttt)){ stop(paste0("The variable",ttt_var," does not contain any
37 ",ttt_ref," value")) }
38
39 data[,ttt := relevel(as.factor(ttt),ref_ttt)] if(is.null(ID_ttt))
40 { data[,ID_ttt := .I] }else{ setnames(data,ID_ttt,"ID_ttt") data[,N := .N,by = ID_ttt]
41 if(any(data$N>1)){ stop("there are",data[N>1,uniqueN(ID_ttt)]," treatment course which
42 have more than one entry in the table. Each row should be an unique treatment") } }
43 return(data) }
44
45
46 adjusted_model = function(data, weights = NULL, covariates = NULL){
47
48 # transform char to factor to_fact <- data[,lapply(.SD,class)] %>% transpose(keep.names =
49 "var") %>% .[V1 == "character",var]
50
51 data[,c(to_fact) := lapply(.SD,factor),.SDcols = to_fact]
52
53 #droplevels facto_vars <- data[,lapply(.SD,class)] %>% transpose(keep.names = "var") %>
54 % .[V1 == "factor",var] data[,c(facto_vars) := lapply(.SD,droplevels),.SDcols = facto_vars]
55
56
57
58
59
60

```

```

1
2
3 # define formula formula <- as.formula(paste0("LDA ~",paste0(c("ttt",covariates),collapse
4 = " + "))
5
6 if(!is.null(covariates)){ # fit fit <- geeglm(formula, data = data, id = ID_ttt, family =
7 gaussian) }else{ fit <- geeglm(LDA ~ ttt, data = data, weights = weights, id = ID_ttt, family =
8 gaussian) }
9
10 fitsummary <- summary(fit) # create table with difference between the two treatments diff
11 <- data.table(ttt = "diff", LDA = fitsummary$coefficients[2,"Estimate"], LDA_var =
12 fitsummary$coefficients[2,"Std.err"]^2, LDA_sup = fitsummary$coefficients[2,"Estimate"] +
13 1.96*fitsummary$coefficients[2,"Std.err"], LDA_inf = fitsummary$coefficients[2,"Estimate"]
14 - 1.96*fitsummary$coefficients[2,"Std.err"], methods = "CC_adjusted")
15
16 # marginal effects: margi_df <- emmeans(fit, "ttt") %>% as.data.table()
17
18 margi_df[,methods := "CC_adjusted"] setnames(margi_df,"emmean","LDA")
19 margi_df[,LDA_inf := LDA - 1.96*SE] margi_df[,LDA_sup := LDA + 1.96*SE]
20 margi_df[,LDA_var := SE^2]
21
22 output <- rbind(diff,margi_df[,.(ttt,LDA,LDA_sup,LDA_inf,LDA_var,methods)])
23
24 return(list(output = output,fit = fit)) }
25
26
27
28
29
30
31 # Not adjusted complete case imputation
32
33
34
35 ```{r setup, include=FALSE}
36
37 export("CC_raw")
38 CC_raw <- function(datain,
39 # data
40 var = "CDAI_fu",
41 # variable measuring effectiveness
42 thres = 10,
43 # threshold for remission or LDA
44 ttt_var = "ttt",
45 ref_ttt = "ttt_ref")
46
47 # variable name containing the treatment
48 {
49 data <- check_data(datain,var,ttt,ref_ttt)
50 # raw proportion
51 raw_prop <- data[!is.na(get(var)),
52 .(LDA = sum(get(var)<=thres)/.N,
53 methods = "CC_raw",
54 N = .N),
55 by = ttt]
56
57
58
59
60

```

```

1
2
3   # calculation of the Standard error
4   raw_prop[,c("LDA_inf","LDA_sup") := lapply(c(-1.96,1.96),function(z)
5   {
6     LDA + z*sqrt(LDA*(1-LDA)/N)
7     })]
8
9   # difference between treatments
10  diff_tmp <- raw_prop[,.(ttt = "diff",
11                          LDA = LDA[ttt == "ttt_1"]-LDA[ttt ==
12 "ttt_ref" ],
13                          methods = methods[1] ,
14                          SE = (sum(1/N))/2 + 1.96*sqrt(sum( LDA*(1-LDA)/N
15 )))]
16
17
18  diff_tmp[,LDA_inf := LDA - SE]
19  diff_tmp[,LDA_sup := LDA + SE]
20
21  # bind outputs
22  output <- rbind(diff_tmp[,.(ttt,LDA,LDA_inf,LDA_sup,methods)],
23                 raw_prop[,-"N"])
24
25  # change name back
26  setnames(output,"ttt",ttt_var)
27  return(output)
28 }
29
30

```

Adjusted complete case imputation

```

33 ```{r setup, include=FALSE} export("CC_adjusted") CC_adjusted = function(datain, var =
34 "CDAI_fu", thres = 10, ttt_var = "ttt", ref_ttt = "ttt_ref", covariates =
35 c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAIO") ) # variable
36 name containing the treatment { data <- check_data(datain,var,ttt_var,ref_ttt) data[,LDA :=
37 get(var) <= thres] output <- adjusted_model(data = data[!is.na(get(var))], covariates =
38 covariates)$output
39
40
41 output[,methods := "CC_adjusted"] # change name back setnames(output,"ttt",ttt_var)
42 return(output) }
43
44
45

```

LOCF imputation

```

46
47
48
49
50 ```{r setup, include=FALSE}
51 export("LOCF")
52
53 LOCF <- function(datain,
54                  var = "CDAI_fu",
55                  var_before = "CDAI_beforefu",
56
57
58
59

```

```

1
2
3         thres = 10,
4         ttt_var = "ttt",
5         ref_ttt = "ttt_ref",
6         covariates =
7 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0")
8 ) {
9   data <- copy(datain)
10
11   data <- check_data(datain, var, ttt_var, ref_ttt)
12   data[is.na(get(var)), c(var) := get(var_before)]
13   data[,LDA := get(var) <= thres]
14
15
16   output <- adjusted_model(data = data,
17                             covariates = covariates)$output
18
19   output[,methods := "LOCF"]
20   # change name back
21   setnames(output, "ttt", ttt_var)
22   return(output)
23 }
24
25
26

```

Lundex imputation

```

27
28
29 ``{r setup, include=FALSE} export("Lundex") Lundex <- function(datain, var = "CDAI_fu",
30 thres = 10, ttt_var = "ttt", ref_ttt = "ttt_ref", treatment_duration = "treatment_duration",
31 stop_var = "stopany", covariates =
32 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0"), boot_num =
33 1000) {
34
35   data <- check_data(datain, var, ttt_var, ref_ttt) data[,LDA := get(var) <= thres] ####
36   bootstrap for SE data[,tmp := 1] # replicated data for bootstrap replicateddata <-
37   data[CJ(tmp = 1, boot = 1:boot_num), on = "tmp", allow.cartesian=TRUE] # sample with
38   replacement for each boot sampled_idx <- replicateddata[,I[sample(1:N, replace = T)], by =
39   boot]$V1 bootstrapdata <- replicateddata[sampled_idx]
40
41
42   # raw proportions raw_prop <- bootstrapdata[!is.na(get(var)), { adjusted_model(data
43   = .SD)$output %>% .[ttt != "diff", (ttt, LDA_raw = LDA) ] }, by = .(boot)]
44
45   # surv analysis for each bootstrapped dataset surv_formula <-
46   as.formula(paste0("Surv(", treatment_duration, ", ", stop_var, ") ~ ttt"))
47
48   surv_coeff <- bootstrapdata[, { temp.km <- survfit(surv_formula, data = .SD) list(surv =
49   summary(temp.km, times = 1)$surv, ttt = gsub("ttt=", "", unique(summary(temp.km)
50   $strata))) }, by = boot]
51
52
53   # LDA: LDA raw * surv coeff tmp_bootstrap <- merge(raw_prop, surv_coeff, by =
54   c("boot", "ttt")) tmp_bootstrap[,LDA := LDA_raw*surv]
55
56
57
58
59
60

```

```

1
2
3 # difference between treatments diff_boot <- tmp_bootstrap[,(ttt = "diff", LDA = LDA[ttt !=
4 ref_ttt] - LDA[ttt == ref_ttt]), by = boot]
5
6 tot_bottstrap <- rbind(diff_boot[,.(ttt, LDA, boot)], tmp_bootstrap[,.(ttt, LDA, boot)])
7
8 # calculate the mean and the SE: output <- tot_bottstrap[,.(LDA = mean(LDA), LDA_sup =
9 quantile(LDA, 0.975), LDA_inf = quantile(LDA, 0.025) ), by = ttt] # change name back
10 output[,methods := "LUNDEX"]
11
12 setnames(output,"ttt",ttt_var) return(output) }
13
14
15
16

```

```

17 # non-responder imputation
18
19

```

```

20 ```{r setup, include=FALSE}
21 export("NRI")
22 NRI = function(datain,
23               var="CDAI_fu",
24               thres = 10,
25               ttt_var = "ttt",
26               ref_ttt = "ttt_ref",
27               covariates =
28 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0")
29 )
30 # variable name containing the treatment
31 {
32   data <- check_data(datain, var, ttt_var, ref_ttt)
33   data[,LDA := get(var) <= thres]
34   data[is.na(LDA),LDA := 0] # missing are non responders
35   output <- adjusted_model(data = data,
36                             covariates = covariates)$output
37
38   # change name back
39   setnames(output, "ttt", ttt_var)
40   output[,methods := "NRI"]
41   return(output)
42 }
43
44
45
46

```

Inverse probability weighting imputation

```

47
48
49 ```{r setup, include=FALSE} export("IPW") IPW <- function(datain, var = "CDAI_fu", thres =
50 10, ttt_var = "ttt", ref_ttt = "ttt_ref", counfounders =
51 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0"), attrition =
52 c("Disease_duration", "concomitantCsDMARD", "Prev_bDMARD3", "CDAI0", "stopreason")) {
53
54   data <- check_data(datain, var, ttt_var, ref_ttt, othervar = c(counfounders, attrition))
55
56
57
58
59
60

```

```

1
2
3 data[,ttt2 := as.numeric(ttt != ref_ttt)] # weight for confounding formula_coeff <-
4 paste0("~",paste0(counfunders,collapse = "+")) function_call <- paste0('IPWT <-
5 ipwpoint( exposure = ttt2 , family = "binomial", link = "logit", numerator = ~ 1,
6 denominator =',formula_coeff,', data = data, trunc = 0.01 )') eval(parse(text = function_call))
7 data$w<- IPW Tipw.weights
8
9
10 # weights for attrition formula_attr <- paste0("~",paste0(attrition,collapse = "+"))
11 data[,MISS := as.numeric(is.na(get(var)))] function_call <- paste0('IPCT <-
12 ipwpoint( exposure = MISS , family = "binomial", link = "logit", numerator = ~ 1,
13 denominator =',formula_attr,', data = data )') eval(parse(text = function_call)) data
14 $w<- IPC Tipw.weights
15
16 dataNoNA <- na.omit(data[,.(ttt,get(var),sw,swc,ID_ttt) %>%
17 setNames(c("ttt",var,"sw","swc","ID_ttt"))]) dataNoNA[,LDA := as.numeric(get(var) <=
18 thres)]
19
20
21 output <- adjusted_model(data = dataNoNA, weights = dataNoNA$w*dataNoNA$swc)
22 $output
23
24 output[,methods := "IPW"]
25
26 # change name back setnames(output,"ttt",ttt_var) return(output)
27
28 }
29
30
31
32 # Confounder-Adjusted Response Rate with Attrition Correction (CARRAC)
33 imputation
34
35
36 ```{r setup, include=FALSE}
37 export("CARRAC")
38 CARRAC <- function(datain,
39                     var = "CDAI_fu",
40                     thres = 10,
41                     ttt_var = "ttt",
42                     ref_ttt = "ttt_ref",
43                     counfunders =
44                     c("Disease_duration","concomitantCsDMARD","Prev_bDMARD3","CDAI0"),
45                     attrition =
46                     c("Disease_duration","concomitantCsDMARD",
47                       "Prev_bDMARD3","CDAI0","stopreason"),
48                     seed = NA) {
49
50   data <- check_data(datain,var,ttt_var,ref_ttt)
51   dataS <- data[,.SD,.SDcols =
52   c("ID_ttt",var,"ttt",union(counfunders,attrition))]
53
54
55
56
57
58
59
60

```

```

1
2
3   impute_data <- mice(
4     dataS,
5     m = 10,
6     method = "pmm",
7     maxit = 5,
8     printFlag = F, seed = seed
9   )
10  # open the data
11  impute_data_complete <- setDT(complete(impute_data, action = "long"))
12  # calculate LDA
13  impute_data_complete[,LDA := get(var) <= thres]
14
15
16  # get LDA and error for each imputation
17  res_mice <- lapply(seq(1:impute_data$m), function(imp){
18
19    adjusted_model(data = impute_data_complete[.imp == imp],
20                  covariates = counfunders)$output
21
22  }) %>% rbindlist()
23
24  res_mice_2 <- lapply(seq(1:impute_data$m), function(imp){
25
26    adjusted_model(data = impute_data_complete[.imp == imp],
27                  covariates = counfunders)$fit
28
29  })
30
31
32  test <- pool(res_mice_2)
33  df_pval <- summary(test) %>% as.data.table()
34  p.output <- df_pval[grepl("ttt", term), p.value]
35
36  # pooling
37  pool_res <- res_mice[,.(
38    LDA_mi = mean(LDA),
39    w = mean(LDA_var),
40    m = .N,
41    b = 1/ (.N-1)*sum( (LDA-mean(LDA))^2 )
42    ),by = ttt]
43
44
45  pool_res[,LDA_var := w + (1+1/m)*b]
46  pool_res[,LDA_sd := sqrt(LDA_var)]
47
48  # mean, 95% CI
49  output <- pool_res[,.(ttt,
50                        LDA_mi,
51                        LDA_mi + 1.96*LDA_sd,
52                        LDA_mi-1.96*LDA_sd) %>%
53                    setNames(c("ttt", "LDA", "LDA_sup", "LDA_inf"))]
54
55  output[,methods := "CARRAC"]
56
57
58
59
60

```



```
1
2
3     output[ttt == "diff",p := p.output]
4
5     # change name back
6     setnames(output,"ttt",ttt_var)
7
8     return(output)
9 }
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

For peer review only

3 - FINAL FIGURES CODE

10/11/2020

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Libraries, Loading data and function

```
library(psych)
library(dplyr)
library(lme4)
library(lmerTest)
library(survival)
library(latticeExtra)
library(Hmisc)
library(mice)
library(car)
library(ggplot2)
library(survminer)
library(xlsx)
library(lubridate)
library(tableone)
library(data.table)
library(stringr)
library(zoo)
library(patchwork) # package to compose multiplots !
library(ggpubr)
library(grid)

rm(list = ls())
setwd(dirname(rstudioapi::getActiveDocumentContext())$path)

load("./3_clean_output/full_workspaces/workspace_1.RData")
load("./3_clean_output/full_workspaces/workspace_2.RData")
load("./3_clean_output/full_workspaces/workspace_3.RData")
```

1. Common theme

```
theme_benoit = function(){
  theme_pubclean()+
    theme(panel.grid.major.x = element_line(linetype = "dotted", colour
= "grey50"),
          panel.grid.major.y = element_line(linetype = "dotted", colour
= "grey50"),
          axis.title.y = element_text(margin = margin(r = .2, unit =
"cm")),
          axis.title.x = element_text(margin = margin(t = .2, unit =
```

```
1  
2  
3 "cm")),  
4 plot.title = element_text(margin = margin(b = .5, unit =  
5 "cm")))  
6 }  
7  
8  
9
```

1. [0] Mini explanation

TC lenght

```
13 BARI_DATA[,time_on_drug_year := time_on_drug/365.25]
```

```
16 p1 <- ggplot(BARI_DATA)+  
17   geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =  
18   1/12)+
```

```
20   scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+  
21   labs(x = "Duration of observation (years)",  
22        y = "Number of TC",  
23        title = "Time of observation for all included TC")+  
24   ylim(-25,NA)+  
25   theme_benoit()
```

p1

```
28  
29 p2 <- ggplot(BARI_DATA)+  
30   geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =  
31   "grey80")+  
32   theme_void()
```

```
34 plot_mini_exploration <- p1 + inset_element(p2,0.01,0.05,0.99,0.2)  
35 plot_mini_exploration
```

Saving plot

```
39 png("./3_clean_output/figures/PLOT_Exploration_TC_duration.png",  
40     width = 7,  
41     height = 5,  
42     units = "in",  
43     res = 300) # opening graphic device  
44 plot_mini_exploration  
45 dev.off() # closing graphic device
```

TC lenght for BARI only

```
49 data_sub <- BARI_DATA[cohort == "BARI"]
```

```
51 p1 <- ggplot(data_sub)+  
52   geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =  
53   1/13, fill = "red3")+
```

```
55   scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
```

```
1
2
3     labs(x = "Duration of observation (years)",
4           y = "Number of TC",
5           title = "A - BARI")+
6     ylim(-11,50)+
7     theme_benoit()
8 p1
9
10
11 p2 <- ggplot(data_sub)+
12     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
13 "grey80")+
14     theme_void()
15
16 plot_mini_exploration_bari <- p1 +
17 inset_element(p2,0.01,0.05,0.99,0.2)
18 plot_mini_exploration_bari
19
20 TC lenght for TNFi only
21
22 data_sub <- BARI_DATA[cohort == "TNFi"]
23
24 p1 <- ggplot(data_sub)+
25     geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
26 1/13, fill = "green2")+
27
28     scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
29     labs(x = "Duration of observation (years)",
30          y = "Number of TC",
31          title = "B - TNFi")+
32     ylim(-11,50)+
33     theme_benoit()
34 p1
35
36
37 p2 <- ggplot(data_sub)+
38     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
39 "grey80")+
40     theme_void()
41
42
43 plot_mini_exploration_tnfi <- p1 +
44 inset_element(p2,0.01,0.05,0.99,0.2)
45 plot_mini_exploration_tnfi
46
47 TC lenght for OMA only
48
49 data_sub <- BARI_DATA[cohort == "OMA"]
50
51 p1 <- ggplot(data_sub)+
52     geom_histogram(aes(x = time_on_drug_year), alpha = .6, binwidth =
53 1/13, fill = "blue2")+
54
55     scale_x_continuous(breaks = c(0,0.5,1,1.5,2,2.5))+
56
57
58
59
60
```

```

1
2
3     labs(x = "Duration of observation (years)",
4           y = "Number of TC",
5           title = "C - OMA")+
6     ylim(-11,50)+
7     theme_benoit()
8 p1
9
10
11 p2 <- ggplot(data_sub)+
12     geom_boxplot(aes(x = time_on_drug_year), alpha = .6, fill =
13 "grey80")+
14     theme_void()
15
16 plot_mini_exploration_oma <- p1 + inset_element(p2,0.01,0.05,0.99,0.2)
17 plot_mini_exploration_oma
18
19 multiplot
20
21 multi_plot <- plot_mini_exploration_bari + plot_mini_exploration_tnfi
22 + plot_mini_exploration_oma
23 multi_plot
24
25 median(BARI_DATA[cohort == "BARI", time_on_drug])
26 median(BARI_DATA[cohort == "TNFi", time_on_drug])
27 median(BARI_DATA[cohort == "OMA", time_on_drug])
28
29 Saving plot
30
31 png("./3_clean_output/figures/
32 PLOT_Exploration_TC_duration_3_groups.png",
33     width = 9,
34     height = 5,
35     units = "in",
36     res = 300) # opening graphic device
37 multi_plot
38 dev.off() # closing graphic device
39
40
41
42
43
44

```

1. [1] Survival analysis

Forest plot BARI vs TNFi + BARI vs OMA

```

45 meanall <- summary(BARI1.adj.mi)$coefficients[1:14,"exp(coef)"]
46 lowerall <- summary(BARI1.adj.mi)$conf.int[1:14,"lower .95"]
47 upperall <- summary(BARI1.adj.mi)$conf.int[1:14,"upper .95"]
48 textall <- c("Treatment (vs BARI)", "Age (decades)", "BMI",
49 "Concomitant csDMARD", "Concomitant glucocorticoid", "CDAI score (10
50 pts)", "Disease duration (decades)", "Current smoker (vs non-smoker)",
51 "Former smoker (vs non-smoker)", "2nd line therapy (vs 1st)", "3rd
52 line therapy (vs 1st)", "4th or later line (vs 1st)", "Female gender",
53 "Seropositivity (RF or ACPA)")
54
55
56
57
58
59

```

```

1
2
3 dfall1 <- data.table(textall, meanall, lowerall, upperall)
4 dfall1[,ttt := "TNFi"]
5
6 meanall <- summary(BARI2.adj.mi)$coefficients[1:14,"exp(coef)"]
7 lowerall <- summary(BARI2.adj.mi)$conf.int[1:14,"lower .95"]
8 upperall <- summary(BARI2.adj.mi)$conf.int[1:14,"upper .95"]
9 dfall2 <- data.table(textall, meanall, lowerall, upperall)
10 dfall2[, ttt := "OMA"]
11
12
13 dfall <- rbind(dfall1,dfall2)
14 dfall$textall <- factor(dfall$textall, levels = rev(textall))
15
16 text_high <- textGrob("\u2192 Reduces \ndrug maintenance",
17 gp=gpar(fontsize=8, fontface="bold"))
18 text_low <- textGrob("\u2190 Improves \ndrug maintenance",
19 gp=gpar(fontsize=8, fontface="bold"))
20
21
22 HR_plot <- ggplot(data=dfall,
23 aes(x = textall,
24 y = meanall,
25 ymin = lowerall,
26 ymax = upperall,
27 color = ttt))+
28 geom_hline(yintercept =1, linetype=2)+
29 geom_point(size=2,position = position_dodge(width = .7))+
30 geom_errorbar(position = position_dodge(width = .7))+
31 labs(x = "",y = "",color = "")+
32 scale_y_log10(breaks=c(0.5,0.6, 0.7, 0.8, 0.9,1,1.2, 1.4, 1.6, 1.8))
33
34 +
35 theme(axis.line.x = element_line(size = 0.5),
36 axis.text = element_text(face = "bold", color = "black"),
37 legend.position="top",
38 legend.key = element_blank(),
39 plot.margin = unit(c(1,3,2,1),"lines"))+
40 coord_flip(clip = "off")+
41 annotation_custom(text_high,
42 xmin=-0.64,xmax=-0.64,ymin=.2,ymax=.2)+
43 annotation_custom(text_low,
44 xmin=-0.64,xmax=-0.64,ymin=-.15,ymax=-.15)+
45 theme_pubclean()+
46 scale_color_manual(breaks = c("OMA","TNFi"),values =
47 c("blue3","green3"),labels = c("OMA","TNFi"))
48

```

HR_plot

Saving the plot in PNG file

```
png("./3_clean_output/figures/PLOT FOREST BARI vs TNFi vs OMA HR.png",
```

```
width = 7,
```

```
1
2
3     height = 5.5,
4     units = "in",
5     res = 300)
6
7 HR_plot
8
9
10 dev.off() # closing graphic device
11
12 BARI vs TNFi
13
14 Non-adjusted Kaplan-Meier curves
15
16 BARI vs TNFi
17
18 BARI1[,time_on_drug_year := time_on_drug/365.25]
19
20 surv_object1 <- Surv(time = BARI1$time_on_drug_year, event =
21 BARI1$stop_DMARD) # indicate time on drug and stop variable
22 fit1 <- survfit(surv_object1 ~ cohort, data = BARI1)
23
24 survplot_1 <- ggsurvplot(fit1, data = BARI1, # plot
25                          pval = T,
26                          pval.method = TRUE,
27                          legend.title = "Groups :",
28                          legend.labs = c("BARI", "TNFi"),
29                          xlab = "Time (years)",
30                          xlim = c(0, 2.5),
31                          censor = FALSE,
32                          title = "Non-adjusted model of drug
33 discontinuation \nby type of treatment",
34                          surv.median.line = "v",
35                          linetype = 1,
36                          size = 1.5,
37                          #palette = c("grey78", "grey10"),
38                          palette = c("red3", "green2"), # pour mettre
39
40 les couleurs
41
42                          ggtheme = theme_benoit(),
43                          risk.table = T)
44
45 values <- summary(fit1)$table[,"median"]
46 df <- data.frame(y = .1,x = values+.2,label =
47 as.character(round(values,2)))
48
49 survplot_1$plot <- survplot_1$plot +
50   geom_text(data = df,aes(x,y,label = label), color = c("red3",
51 "green2"), size = 5)
52
53
54
55 print(survplot_1)
56
57
58
59
60
```

1
2
3 Saving surplot

4
5 png("./3_clean_output/figures/PLOT BARI vs TNFi curves non adjusted
6 COLOR.png",
7 width = 7,
8 height = 7, units = "in",
9 res = 300) # opening graphic device
10 survplot_1

11
12 dev.off() # closing graphic device

13
14 Saving the plot curv object for Lilly

15
16 plot_BARI_vs_TNFi_data <- survplot_1\$data.survplot
17 write.xlsx(plot_BARI_vs_TNFi_data, file =
18 "./3_clean_output/Lilly_curves_excel/plot_BARI_vs_TNFi_data_non_adjust
19 ed.xlsx", row.names = F)

20
21 **Home-made attempt to obtain adjusted curves based on imputed data**

22 dummy_cox_impute1 <- mice::complete(imputed_data1, "long", include =
23 T)
24 dummy_cox_impute1 <- dummy_cox_impute1[dummy_cox_impute1\$.imp != 0,]
25 dummy_cox_impute1\$time_on_drug_year <-
26 dummy_cox_impute1\$time_on_drug/365.25

27
28
29 BARI_fit1 <- survfit(coxph(Surv(time = time_on_drug_year, event =
30 stop_DMARD) ~ cohort+

31 I(age_base/10)+
32 bmi_base+
33 TC_with_csDMARD+
34 PREDNISON_STEROID+
35 CDAI0+
36 I(disease_duration_base_years/10)+
37 C(smoker_base, base=3)+
38 line_of_therapy+
39 gender+
40 seropositivity_base+
41 cluster(patient_id)+
42 strata(cohort), dummy_cox_impute1), data =

43
44 dummy_cox_impute1)

45
46
47
48 survplot_1_adj <- ggsurvplot(BARI_fit1, data = dummy_cox_impute1,
49 variable = "cohort",

50 xlab = "Time (years)",
51 title = "A - BARI vs TNFi",
52 legend.title = "Groups :",
53 legend.labs = c("BARI", "TNFi"),
54 censor = FALSE,
55 xlim = c(0, 2.5),
56
57
58
59


```
1
2
3         surv.median.line = "v",
4         linetype = 1,
5         size = 1.5,
6         ggtheme = theme_benoit(),
7         # palette = c("grey78", "grey10")
8         palette = c("red2", "green3" )+
9
10        labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,\n
11        concomitant glucocorticoid, baseline CDAI, disease duration
12        (decades),\n smoking status, line of therapy, gender, seropositivity")
13
14        # adding days label
15        values <- summary(BARI_fit1)$table[,"median"]
16        df <- data.frame(y = .1,x = values+.1,label =
17        as.character(paste(round(values*365.25,2), "\n days")))
18        df[1,2] <- 1.82
19
20        survplot_1_adj$plot$labels$y <- "Proportion still on drug" # to change
21        the label
22
23        survplot_1_adj$plot <- survplot_1_adj$plot +
24        geom_text(data = df,aes(x,y,label = label), color = c("red3",
25        "green3"), size = 5)
26
27        # adding HR et p val label
28        HR <- data.frame(y = 0.1, x = 0.5, label = paste("HR =",
29        round(exp(BARI1.adj.mi$coefficients[1]), 2), "\n", "p =",
30        round(summary(BARI1.adj.mi)$coefficients[1,"Pr(>|z|)"], 4)  ) )
31
32        survplot_1_adj$plot <- survplot_1_adj$plot +
33        geom_text(data = HR,aes(x,y,label = label) , size = 5)
34
35        # final print
36        survplot_1_adj
37
38        Saving the survival plot in PNG file
39
40        png("./3_clean_output/figures/PLOT BARI vs TNFi curves adjusted.png",
41
42        width = 7,
43        height = 5,
44        units = "in",
45        res = 300)
46
47        survplot_1_adj
48
49        dev.off() # closing graphic device
```

Sensitivity analysis (RiskRegression Package)

First plot to get the difference in average treatment effect in percentage

```

1
2
3 dt.out$time_years <- dt.out$time/365.25
4
5 plot.ate.diff <- ggplot(dt.out[type == "meanRisk"], aes(x =
6 time_years, group = level))+
7   geom_vline(xintercept = 1, linetype = 2, size = 1, color = "grey20")
8 +
9   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
10 0.4)+
11   geom_line(aes(y = estimate, color = level), size = 1)+
12
13   scale_colour_manual(values = c("red2","green3"))+
14   scale_fill_manual(values = c("red2","green3"))+
15   theme_minimal() + theme(legend.spacing.x = unit(0.2, 'cm'),
16 legend.position="top" )+
17   scale_x_continuous(breaks=seq(0,2.5,0.25)) +
18   scale_y_continuous(labels = scales::percent, limits = c(0,0.65))+
19
20   xlab("Years since initiation of treatment")+
21   ylab("Absolute Risk of treatment discontinuation (%)")+
22   labs(colour="Groups:", fill = "Groups:", title = "A - BARI vs TNFi")
23 +
24   labs(group = "Groups:")+
25   theme_benoit()+
26   theme(axis.title.x = element_text(margin = margin(t = .3,unit =
27 "cm")),
28 axis.title.y = element_text(margin = margin(r = .3,unit =
29 "cm")))
30
31
32
33 plot.ate.diff
34
35 Saving Plot
36
37 png("./3_clean_output/figures/PLOT BARI vs TNFi curves AIPTW.png",
38 width = 1300, height = 650, res = 120) # opening graphic device
39
40 plot.ate.diff
41
42 dev.off() # closing graphic device
43
44 Second plot to get the ratio in average treatment effect
45
46 plot.ate.ratio <- ggplot(dt.out[type == "ratioRisk"], aes(x = time,
47 group = level))+
48   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
49 0.4)+
50   geom_line(aes(y = estimate, color = level), size = 2)+
51
52   theme_benoit()+
53   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
54   scale_x_continuous(breaks=seq(0,500,50))+
55
56
57
58
59

```

```
1
2
3     scale_y_continuous(limits = c(0.9,4.5))+
4
5     xlab("Days since intiation of treatment")+
6     ylab("Ratio in Average Treatment Effect")+
7     labs(colour="treatment", fill = "treatment")
8
```

```
9
10 plot.ate.ratio
```

11 BARI vs OMA

12 Non-adjusted Kaplan-Meier curves

13 BARI vs OMA

```
14 BARI2[,time_on_drug_year := time_on_drug/365.25]
```

```
15
16 surv_object2 <- Surv(time = BARI2$time_on_drug_year, event =
17 BARI2$stop_DMARD)
18 fit2 <- survfit(surv_object2 ~ cohort, data = BARI2) # this function
19 creates the data for Kaplan Meyer
20 survplot_2 <- ggsurvplot(fit2, data = BARI2, # plot
21     pval = T,
22     pval.method = TRUE,
23     legend.title = "Groups :",
24     legend.labs = c("BARI", "OMA"),
25     xlab = "Time (days)",
26     xlim = c(0, 2.5),
27     censor = FALSE,
28     title = "Non-adjusted model of drug discontinuation by type
29 of treatment",
30     surv.median.line = "v",
31     linetype = 1,
32     size = 1.5,
33     ggtheme = theme_benoit(),
34     #palette = c("grey78", "grey50"),
35     palette = c("red3", "blue2"), # to put colors
36     risk.table = T)
37
```

```
38 survplot_2
```

39 Saving surplot

```
40 png("./3_clean_output/PLOT BARI vs OMA curves non adjusted COLOR.png",
41 width = 1000, height = 600, res = 100) # opening graphic device
```

```
42 survplot_2
```

```
43 dev.off() # closing graphic device
```

44 Saving the plot curv object for Lilly

```

1
2
3 plot_BARI_vs_OMA_data <- survplot_2$data.survplot
4 write.xlsx(plot_BARI_vs_OMA_data, file =
5 ".\3_clean_output\Lilly_curves_excel\plot_BARI_vs_OMA_data_non_adjus
6 t.d.xlsx", row.names = F)
7

```

Home-made attempt to obtain adjusted cuves based on imputed data :

```

8
9 dummy_cox_impute2 <- mice::complete(imputed_data2, "long", include =
10 T)
11 dummy_cox_impute2 <- dummy_cox_impute2[dummy_cox_impute2$.imp != 0,]
12 dummy_cox_impute2$time_on_drug_year <-
13 dummy_cox_impute2$time_on_drug/365.25
14
15
16 BARI_fit2 <- survfit(coxph(Surv(time = time_on_drug_year, event =
17 stop_DMARD) ~ cohort+
18
19 I(age_base/10)+
20 bmi_base+
21 TC_with_csDMARD+
22 PRÉDNISON_STEROID+
23 CDAI0+
24 I(disease_duration_base_years/10)+
25 C(smoker_base, base=3)+
26 line_of_therapy+
27 gender+
28 seropositivity_base+
29 cluster(patient_id)+
30 strata(cohort),dummy_cox_impute2), data =
31 dummy_cox_impute2)
32
33 survplot_2_adj <- ggsvplot(BARI_fit2, data = dummy_cox_impute2,
34 variable = "cohort",
35 xlab = "Time (years)",
36 title = "B - BARI vs OMA",
37 legend.title = "Groups :",
38 legend.labs = c("BARI", "OMA"),
39 censor = FALSE,
40 xlim = c(0, 2.5),
41 surv.median.line = "v",
42 linetype = 1,
43 size = 1.5,
44 ggtheme = theme_benoit(),
45 # palette = c("grey78", "grey50")
46 palette = c("red2", "blue3") # to change colors
47 )+
48
49 labs(caption = "Adjusted for : age, BMI, concomitant csDMARD, \n
50 concomitant glucocorticoid, baseline CDAI, disease duration
51 (decades),\n smoking status, line of therapy, gender, seropositivity")
52
53 # adding Days label
54 values <- summary(BARI_fit2)$table[,"median"]
55 df <- data.frame(y = .1,x = values+.1,label =
56
57
58
59
60

```

```
1
2
3 as.character(paste(round(values*365.25,2), "\n days"))
4 df[1,2] <- 1.82
5
6 survplot_2_adj$plot$labels$y <- "Proportion still on drug" # to change
7 the label
8
9
10 survplot_2_adj$plot <- survplot_2_adj$plot +
11   geom_text(data = df,aes(x,y,label = label), color = c("red3",
12 "blue2"), size = 5)
13
14 # adding HR et pval label
15 HR <- data.frame(y = 0.1, x = 0.5, label = paste("HR =",
16 round(exp(BARI2.adj.mi$coefficients[1]), 2), "\n", "p =",
17 round(summary(BARI2.adj.mi)$coefficients[1,"Pr(>|z|)"], 4) ) )
18
19 survplot_2_adj$plot <- survplot_2_adj$plot +
20   geom_text(data = HR,aes(x,y,label = label) , size = 5)
21
22
23
24
25 # final print
26 survplot_2_adj
27
28 summary(BARI_fit2, times = 1) # to see detailed surv probabilities at
29 given timepoints
30
31 Saving the survival plot in PNG file
32
33 png("./3_clean_output/PLOT BARI vs OMA curves adjusted.png", width =
34 1000, height = 600, res = 100) # opening graphic device
35
36 survplot_2_adj
37
38 dev.off() # closing graphic device
39
40
41 Sensitivity analysis (RiskRegression package)
42
43 First plot to get the difference in average treatment effect in percentage
44
45 dt.out2$time_years <- dt.out2$time/365.25
46
47 plot.ate.diff2 <- ggplot(dt.out2[type == "meanRisk"], aes(x =
48 time_years, group = level))+
49   geom_vline(xintercept = 1, linetype = 2, size = 1, color = "grey20")
50 +
51   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
52 0.3)+
53   geom_line(aes(y = estimate, color = level), size = 1)+
54
55 theme_benoit() + theme(legend.spacing.x = unit(0.2, 'cm'),
56
57
58
59
```

```

1
2
3 legend.position="top" )+
4   scale_x_continuous(breaks=seq(0,2.5,0.25)) +
5   scale_y_continuous(labels = scales::percent, limits = c(0,0.65))+
6
7   xlab("Years since initiation of treatment")+
8   ylab("Absolute Risk of treatment discontinuation (%)")+
9   labs(colour="Groups:", fill = "Groups:", title = "B - BARI vs OMA")+
10  labs(group = "Groups:")+
11
12
13   scale_colour_manual(values = c("red2","blue3"))+
14   scale_fill_manual(values = c("red2","blue3"))

```

```
15
16 plot.ate.diff2
```

```
17
18 Saving Plot
```

```
19
20 png("./3_clean_output/PLOT BARI vs OMA curves AIPTW.png", width =
21 1300, height = 650, res = 120) # opening graphic device
22
```

```
23
24 plot.ate.diff2
```

```
25
26 dev.off() # closing graphic device
```

```
27
28 Second plot to get the ratio in average treatment effect
```

```
29
30 plot.ate.ratio2 <- ggplot(dt.out2[type == "ratioRisk"], aes(x = time,
31 group = level))+
32   geom_ribbon(aes(ymin = lower, ymax = upper, fill = level), alpha =
33 0.3)+
34   geom_line(aes(y = estimate, color = level), size = 1)+
35
36   theme_benoit()+
37   theme(legend.spacing.x = unit(0.2, 'cm'), legend.position="top")+
38   scale_x_continuous(breaks=seq(100,400,50))+
39   scale_y_continuous(limits = c(0.8,3))+
40
41   xlab("Days since initiation of treatment")+
42   ylab("Ratio in Average Treatment Effect")+
43   labs(colour="treatment", fill = "treatment")
44
```

```
45
46 plot.ate.ratio2
```

47 Multipanel plots

48
49 To update using patchwork

50
51 For the paper Non adjusted curves

52
53 # Creating list object

```
54
55 plots <- list()
56
57
58
59
60
```

```
1
2
3 plots[[1]] <- survplot_1
4 plots[[2]] <- survplot_2
5
6 # Nice function
7
8 multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
9
10 Option 2 putting all data on one panel Kaplan Meier
11
12 BARI vs TNFi vs OMA
13
14 BARI_DATA[,time_on_drug_year := time_on_drug/365.25]
15
16
17 surv_object3 <- Surv(time = BARI_DATA$time_on_drug_year, event =
18 BARI_DATA$stop_DMARD)
19 fit3 <- survfit(surv_object3 ~ cohort, data = BARI_DATA) # this
20 function creates the data for Kaplan Meyer
21 survplot_3 <- ggsurvplot(fit3, data = BARI_DATA, # plot
22 pval = F,
23 pval.method = F,
24 legend.title = "Groups",
25 legend.labs = c("BARI", "TNFi", "OMA"),
26 xlab = "Time (years)",
27 xlim = c(0, 2.5),
28 censor = FALSE,
29 # title = "Non-adjusted drug discontinuation by type of
30 treatment (Kaplan-Meier)",
31 surv.median.line = "v",
32 linetype = 1,
33 size = 1.5,
34 ggtheme = theme_benoit(),
35 palette = c("red3", "green2", "blue2"), # to put colors
36 risk.table = T)
37
38
39 values <- summary(fit3)$table[, "median"]
40 df <- data.frame(y = .1, x = values+.1, label =
41 as.character(paste(round(values*365.25,2), "\n days")))
42 df[3,2] <- 1.72
43
44
45 survplot_3$plot <- survplot_3$plot +
46 geom_text(data = df, aes(x,y,label = label), color = c("red3",
47 "green2", "blue2"), size = 5)
48
49 survplot_3$plot$labels$y <- "Proportion still on drug" # to change the
50 label
51 survplot_3
52
53 Saving survplot
54
55 png("./3_clean_output/figures/PLOT BARI vs TNFi vs OMA curves non
56 adjusted COLOR.png", width = 800, height = 600, res = 100) # opening
57
58
59
60
```

```
1
2
3     graphic device
4
5     survplot_3
6
7     dev.off() # closing graphic device
8
9     Adjusted curves
10
11    # Creating list object
12
13    plots <- list()
14    plots[[1]] <- survplot_1_adj
15    plots[[2]] <- survplot_2_adj
16
17    # Nice function
18
19
20    multi_plot_cox <- arrange_ggsurvplots(plots, print = T, ncol = 2)
21
22    png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA curves
23    adjusted COLOR.png", width = 1000, height = 600, res = 100) # opening
24    graphic device
25
26    multi_plot_cox
27
28
29    dev.off() # closing graphic device
30
31    All curves
32
33    # Creating list object
34
35    plots <- list()
36    plots[[1]] <- survplot_1
37    plots[[3]] <- survplot_2
38    plots[[2]] <- survplot_1_adj
39    plots[[4]] <- survplot_2_adj
40
41    # Nice function
42
43
44    multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2, nrow =
45    2)
46
47    # but does not display properly now.. :(
48
49    AIPTW absolute risk of treatment discontinuation biplot
50
51    plot.ate.diff + plot.ate.diff2
52
53    png("./3_clean_output/figures/BIPLOT BARI vs TNFi vs OMA AIPTW curves
54    adjusted COLOR.png", width = 1000, height = 600, res = 100) # opening
55    graphic device
56
57
58
59
```



```
1
2
3 plot.ate.diff + plot.ate.diff2
4
5 dev.off() # closing graphic device
6
```

7 Diagnostic multipanel plots

8 Asked by Lilly statistician to show balance in this analysis.

```
9
10
11 pscore_plot <- ggplot(test.data, aes(x = pscores, color = cohort, fill
12 = cohort)) +
13   geom_density(alpha = .47) +
14
15   theme_minimal()+
16   theme(axis.ticks.y = element_blank(),
17         panel.grid.minor = element_blank(),
18         legend.title = element_blank(),
19         text = element_text(size = 16),
20         axis.title.x = element_text(hjust = 0.2, size = 16))+
21
22   scale_colour_manual(values = c("red2","green3"))+
23   scale_fill_manual(values = c("red2","green3"))+
24
25   xlab("Probability of being assigned BARI or TNFi") +
26   ylab("Density") +
27   labs(title = "A1 - BARI vs TNFi")
28
29
```

```
30
31 pscore_plot # overlap
```

```
32
33 pscore_plot2 <- ggplot(test.data2, aes(x = pscores, color = cohort,
34 fill = cohort)) +
35   geom_density(alpha = .47) +
36
37   theme_minimal()+
38   theme(axis.ticks.y = element_blank(),
39         panel.grid.minor = element_blank(),
40         legend.title = element_blank(),
41         text = element_text(size = 16),
42         axis.title.x = element_text(hjust = 0.2, size = 16))+
43
44   scale_colour_manual(values = c("red2","blue3"))+
45   scale_fill_manual(values = c("red2","blue3"))+
46
47   xlab("Probability of being assigned BARI or OMA") +
48   ylab("Density") +
49   labs(title = "A2 - BARI vs OMA")
50
```

```
51
52 pscore_plot2
53 # Good overlap
54
55
56
57
58
59
```

```

1
2
3   library(cobalt)
4
5   # BARI vs TNFi
6   B1 <- love.plot(COVS, treat = test.data$cohort, weights =
7   test.data$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
8   var.order = "adjusted", title = "B1 - BARI vs TNFi", color =
9   c("#FD8D3C", "#08306B"), themes = theme_pubclean() )
10
11
12  # BARI vs OMA
13  B2 <- love.plot(COVS_2, treat = test.data2$cohort, weights =
14  test.data2$weights, stats = c("mean.diffs"), thresholds = c(m = .1),
15  var.order = "adjusted", title = "B2 - BARI vs OMA", color =
16  c("#FD8D3C", "#08306B") , themes = theme_pubclean() )
17
18  one <- ( pscore_plot + B1)
19
20  two <- ( pscore_plot2 + B2 )
21
22  png("./3_clean_output/figures/AIPTW diagnostc COLOR.png", width =
23  1300, height = 900, res = 100) # opening graphic device
24
25  one / two
26
27
28  dev.off() # closing graphic device
29
30

```

1. [3] Fist line analysis

Non-adjusted Survival curves

BARI vs TNFi

```

35
36
37  BARI_first1 <- copy(BARI_first[cohort %in% c("BARI", "TNFi")])
38
39
40  surv_object3 <- Surv(time = BARI_first1$time_on_drug, event =
41  BARI_first1$stop_DMARD) # indiqate stop variable and time_on_drug
42  summary(coxph(surv_object3 ~ cohort, data = BARI_first1))
43  fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first1) # function
44  which creates Kaplan-meier data
45  survplot_first1 <- ggsurvplot(fit3, data = BARI_first1, # plot
46  pval = T,
47  pval.method = TRUE,
48  legend.title = "Groups :",
49  legend.labs = c("BARI", "TFNi"),
50  xlab = "Time (days)",
51  xlim = c(0, 700),
52  censor = FALSE,
53  title = "A - BARI vs TNFi",
54  surv.median.line = "v",
55  linetype = 1,
56
57
58
59

```

```
1
2
3     size = 1.5,
4     ggtheme = theme_benoit(),
5     # palette = c("grey78", "grey50"),
6     palette = c("red2", "green3"), # to get colors
7     risk.table = T
8   )
9
10  survplot_first1$plot$labels$y <- "Proportion still on drug" # to
11  change the label
12  survplot_first1
13
14
15  rm(surv_object3, fit3)
16
17  saving plot curves
18
19  png("./3_clean_output/PLOT BARI vs TNFi first line curves non adjusted
20  COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
21  device
22
23  survplot_first1
24
25  dev.off() # closing graphic device
26
27  BARI vs OMA
28
29  BARI_first2 <- BARI_first[line_of_therapy == "1st" & cohort %in%
30  c("BARI", "OMA")] # selection des TC TNFi
31
32
33  surv_object3 <- Surv(time = BARI_first2$time_on_drug, event =
34  BARI_first2$stop_DMARD) # indiquer stop variable and time_on_drug
35  summary(coxph(surv_object3 ~ cohort, data = BARI_first2))
36  fit3 <- survfit(surv_object3 ~ cohort, data = BARI_first2) # fonction
37  which creates Kaplan-meier data
38  survplot_first2 <- ggsurvplot(fit3, data = BARI_first2, # plot
39    pval = T,
40    pval.method = TRUE,
41    legend.title = "Groups :",
42    legend.labs = c("BARI", "OMA"),
43    xlab = "Time (days)",
44    xlim = c(0, 700),
45    censor = FALSE,
46    title = "B - BARI vs OMA",
47    surv.median.line = "v",
48    linetype = 1,
49    size = 1.5,
50    ggtheme = theme_benoit(),
51    # palette = c("grey78", "grey50", "grey10"),
52    palette = c("red2", "blue3"), # to get colors
53    risk.table = T
54  )
55
56
57
58
59
60
```

```

1
2
3
4   survplot_first2$plot$labels$y <- "Proportion still on drug" # to
5   change the label
6   survplot_first2
7
8
9   rm(surv_object3, fit3)
10
11  saving plot curves
12
13  png("./3_clean_output/PLOT BARI vs OMA first line curves non adjusted
14  COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
15  device
16
17  survplot_first2
18
19  dev.off() # closing graphic device
20
21  Adjusted curves with imputed data (BARI vs TNFi)
22  dummy_cox_impute_first1 <- mice::complete(imputed_data1_first, "long",
23  include = T)
24  dummy_cox_impute_first1 <-
25  dummy_cox_impute_first1[dummy_cox_impute_first1$.imp != 0,]
26
27  BARI_first1_fit <- survfit(coxph(Surv(time = time_on_drug, event =
28  stop_DMARD) ~ cohort+
29
30      I(age_base/10)+
31      bmi_base+
32      concomitant_csDMARD+
33      PREDNISON_STEROID+
34      CDAI0+
35      I(disease_duration_base_years/10)+
36      C(smoker_base, base=3)+
37      line_of_therapy+
38      gender+
39      seropositivity_base+
40      cluster(patient_id)+
41      strata(cohort), dummy_cox_impute_first1),
42  data = dummy_cox_impute_first1)
43
44
45
46  survplot_first1_adj <- ggsurvplot(BARI_first1_fit, data =
47  dummy_cox_impute_first1, variable = "cohort",
48      xlab = "Time (days)",
49      title = "Multivariable Cox model of drug discontinuation by
50  type of treatment - 1st line vs 1st line",
51      legend.title = "Groups :",
52      legend.labs = c("Baricitinib", "TNFi"),
53      censor = FALSE,
54      xlim = c(0, 700),
55      surv.median.line = "v",
56
57
58
59
60

```

```
1
2
3     linetype = 1,
4     size = 1.5,
5     ggtheme = theme_minimal(),
6     #palette = c("grey78", "grey10")
7     palette = c("red2", "green3"), # to get colors
8   )
9
10  survplot_first1_adj <- survplot_first1_adj +
11    labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
12    concomitant glucocorticoid, baseline CDAI, disease duration (decades),
13    smoking status, line of therapy, gender, seropositivity")
14
15
16  survplot_first1_adj
17  table(BARI_first1$cohort)
18  rm(dummy_cox_impute_first1, BARI_first1_fit)
19
20  Saving the survival plot in PNG file
21
22  png("./3_clean_output/PLOT BARI vs TNFi first line curves adjusted
23  COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
24  device
25
26  survplot_first1_adj
27
28
29  dev.off() # closing graphic device
30
31  Adjusted curves with imputed data (BARI vs OMA)
32  dummy_cox_impute_first2 <- mice::complete(imputed_data2_first, "long",
33  include = T)
34  dummy_cox_impute_first2 <-
35  dummy_cox_impute_first2[dummy_cox_impute_first2$.imp != 0,]
36
37  BARI_first2_fit <- survfit(coxph(Surv(time = time_on_drug, event =
38  stop_DMARD) ~ cohort+
39      I(age_base/10)+
40      bmi_base+
41      concomitant_csDMARD+
42      PREDNISON_STEROID+
43      CDAI0+
44      I(disease_duration_base_years/10)+
45      C(smoker_base, base=3)+
46      line_of_therapy+
47      gender+
48      seropositivity_base+
49      cluster(patient_id)+
50      strata(cohort), dummy_cox_impute_first2),
51  data = dummy_cox_impute_first2)
52
53
54
55  survplot_first2_adj <- ggsurvplot(BARI_first2_fit, data =
```

```
1
2
3 dummy_cox_impute_first2, variable = "cohort",
4     xlab = "Time (days)",
5     title = "Multivariable Cox model of drug discontinuation by
6 type of treatment - 1st line vs 1st line",
7     legend.title = "Groups :",
8     legend.labs = c("Baricitinib", "OMA bDMARDs"),
9     censor = FALSE,
10    xlim = c(0, 700),
11    surv.median.line = "v",
12    linetype = 1,
13    size = 1.5,
14    ggtheme = theme_minimal(),
15    #palette = c("grey78", "grey50")
16    palette = c("red2", "blue3"), # to get colors
17    )
18
19
```

```
20 survplot_first2_adj <- survplot_first2_adj +
21     labs(caption = "Adjusted for : age, BMI, concomitant csDMARD,
22 concomitant glucocorticoid, baseline CDAI, disease duration (decades),
23 smoking status, line of therapy, gender, seropositivity")
24
```

```
25
26 survplot_first2_adj
27 table(BARI_first2$cohort)
28 rm(dummy_cox_impute_first2, BARI_first2_fit)
29
```

30 Saving the survival plot in PNG file

```
31 png("./3_clean_output/PLOT BARI vs OMA first line curves adjusted
32 COLOR.png", width = 1000, height = 600, res = 100) # opening graphic
33 device
34
```

```
35
36 survplot_first2_adj
37
```

```
38 dev.off() # closing graphic device
39
```

40 Multipanel plots

41 Non adjusted curves

```
42
43 plots <- list()
44 plots[[1]] <- survplot_first1
45 plots[[2]] <- survplot_first2
46
47
```

48 # Nice function

```
49
50 multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
51
```

```
52
53 png("./3_clean_output/figures/BILOT BARI vs TNFi vs OMA 1st Line
54 curves non-adjusted COLOR.png", width = 1000, height = 600, res = 100)
55 # opening graphic device
56
57
58
59
```

```
1
2
3
4 multi_plot <- arrange_ggsurvplots(plots, print = T, ncol = 2)
5
6 dev.off() # closing graphic device
7
8 All in one BARI vs TNFi vs OMA
9
10 BARI_first[,time_on_drug_year := time_on_drug/365.25]
11
12 surv_object4 <- Surv(time = BARI_first$time_on_drug_year, event =
13 BARI_first$stop_DMARD)
14 fit4 <- survfit(surv_object4 ~ cohort, data = BARI_first) # this
15 function creates the data for Kaplan Meyer
16
17
18 survplot_4 <- ggsurvplot(fit4, data = BARI_first, # plot
19 pval = F,
20 pval.method = F,
21 legend.title = "Groups",
22 legend.labs = c("BARI", "TNFi", "OMA"),
23 xlab = "Time (years)",
24 xlim = c(0, 2.5),
25 censor = FALSE,
26 # title = "Non-adjusted drug discontinuation by type of
27 treatment (Kaplan-Meier)",
28 surv.median.line = "v",
29 linetype = 1,
30 size = 1.5,
31 ggtheme = theme_benoit(),
32 palette = c("red3", "green2", "blue2"), # to put colors
33 risk.table = T)
34
35
36 values <- summary(fit4)$table[, "median"]
37 df <- data.frame(y = .2, x = values+.2, label =
38 as.character(paste(round(values*365.25, 2), "\n days")))
39 df <- df[2,]
40
41 survplot_4$plot <- survplot_4$plot +
42   geom_text(data = df, aes(x, y, label = label), color = c("green2"),
43   size = 5)
44
45
46 survplot_4$plot$labels$y <- "Proportion still on drug" # to change the
47 label
48 survplot_4
49
50 Saving survplot
51
52 png("./3_clean_output/figures/PLOT BARI vs TNFi vs OMA first curves
53 non adjusted COLOR.png", width = 800, height = 600, res = 100) #
54 opening graphic device
55
56
57
58
59
60
```

```

1
2
3     survplot_4
4
5     dev.off() # closing graphic device
6
7

```

1. [4] LACK of EFFICACY and ADVERSE EVENTS

Analysis by stop_reasons in competing risk

(BARI vs TNFi)

Cumulative incidence function

```

16     ci_long$time_months <- ci_long$time/365.25*12
17
18     plot2 <- ggplot(data = ci_long, aes(x = time_months,
19                                         y = value,
20                                         linetype = variable ,
21                                         col = cohort )) +
22
23         geom_line(size = 0.75)+
24
25         scale_color_manual(breaks = c(0,1),
26                             values = c("red3","green2"),
27                             labels = c("BARI","TNFi"))+
28         scale_linetype_manual(breaks = c("CI.1","CI.2"),
29                               values = c("solid","dashed", "dotted"),
30                               labels = c("Adverse Event","Ineffectiveness"))
31 + # not showing the "other category"
32     scale_x_continuous(name = "Time (months)",
33                       breaks = c(0,3,6,9,12),
34                       limits = c(0,12)) +
35     scale_y_continuous(name = "Cumulative incidence", limits = c(0,0.4))
36 +
37     theme_benoit()+
38     theme(legend.box = "horizontal",
39           legend.position = c(0.05,1),
40           legend.justification = c(0,1),
41           legend.background = element_blank(),
42           legend.key = element_blank(), #no legend key background
43           legend.key.width = grid::unit(2, "lines"))+ # longer line in
44     legend, to see properly the dashed
45     labs(color = "", linetype = "", title = "A - BARI vs TNFi")
46
47
48     plot2
49
50     ggsave(filename = "PLOT BARI vs TNFi cumulative incidence.png",plot =
51     plot2, path = "./3_clean_output/", device = "png", width = 829, height
52     = 550, units = "px", scale = 3.2)
53
54
55
56
57
58
59
60

```


(BARI vs OMA)*Cumulative incidence function*

```

1
2
3
4
5
6 ci_long_2$time_months <- ci_long_2$time/365.25*12
7
8
9 plot3 <- ggplot(data = ci_long_2, aes(x = time_months,
10                                     y = value,
11                                     linetype = variable ,
12                                     col = cohort )) +
13
14   geom_line(size = 0.75)+
15
16   scale_color_manual(breaks = c(0,1),
17                      values = c("red3","blue2"),
18                      labels = c("BARI","OMA"))+
19   scale_linetype_manual(breaks = c("CI.1","CI.2"), # not showing the
20 "other" category
21                        values = c("solid","dashed", "dotted"),
22                        labels = c("Adverse Event","Ineffectiveness"))
23
24 +
25   scale_x_continuous(name = "Time (months)",
26                      breaks = c(0,3,6,9,12),
27                      limits = c(0,12)) +
28   scale_y_continuous(name = "Cumulative incidence", limits = c(0,0.4))
29
30 +
31   theme_benoit()+
32   theme(legend.box = "horizontal",
33         legend.position = c(0.05,1),
34         legend.justification = c(0,1),
35         legend.background = element_blank(),
36         legend.key = element_blank(), #no legend key background
37         legend.key.width = grid::unit(2, "lines"))+ # longer line in
38 legend, to see properly the dashed
39   labs(color = "", linetype = "", title = "B - BARI vs OMA")
40
41 plot3
42
43 ggsave(filename = "./3_clean_output/figures/PLOT BARI vs OMA
44 cumulative incidence.png", plot3, height = 4, width = 6, units =
45 "in",dpi = 300)

```

Multipanel

```

46
47 plot2_3 <- plot2 + plot3
48
49
50
51
52
53
54
55
56
57
58
59
60

```

```

59 ggsave(filename = "./3_clean_output/figures/PLOT BARI vs TNFi and BARI
60 vs OMA cumulative incidence.png", plot2_3, height = 4, width = 8,
61 units = "in",dpi = 300)

```

1. [6] LDA - REM

Exploration

See all available raw CDAI measures : :)

```

BARI_long[, group := "non-BARI"]
BARI_long[drug == "BIOLOGIC_BARICITINIB", group := "BARI"]

plot_data <- copy(BARI_long[!is.na(TC_id) & TC_id %in%
BARI_DATA$TC_id])
plot_data <- merge(plot_data, BARI_DATA[,.(TC_id, cohort)], by =
"TC_id")

CDAI_plot <- ggplot(data = plot_data,
                    aes(x = time, y = CDAI, fill = cohort) )+

  annotate("rect",xmin = 0.875,xmax = 1.125,
          ymin = -1,ymax = 60,alpha = .5,fill = "grey80")+
  annotate("rect",xmin = -.05,xmax = 0+1.5/12,
          ymin = -1,ymax = 60,alpha = .5,fill = "grey80")+

  geom_point(data = plot_data[cohort != "BARI"], alpha = 0.2, size =
2, shape = 21, position = position_jitter(width = 0.02, seed = 123) )+
  geom_point(data = plot_data[cohort == "BARI"], alpha = 0.25, size =
2, shape = 21, position = position_jitter(width = 0.02, seed = 123) )+

  #geom_jitter(width = 0.01, height = 0.01, data = plot_data[cohort !=
"BARI"], alpha = 0.2, size = 2, shape = 21, show.legend = F )+
  #geom_jitter(width = 0.01, height = 0.01, data = plot_data[cohort ==
"BARI"], alpha = 0.25, size = 2, shape = 21 , show.legend = F, )+

  geom_smooth(alpha = 0.1, size = 1, aes(color = cohort), show.legend
= F)+

  coord_cartesian(xlim = c(0,2.5))+
  labs(title = "CDAI across time type of treatment (all TC)",
       x = "Time (years since TC initiation)",
       y = "CDAI score",
       color = "",
       fill = "")+
  theme_benoit()+
  theme(legend.position = c(1,1),
        legend.justification = c(1,1))+
  guides(color = guide_legend(override.aes = list(linetype = NA,size =
3)))

CDAI_plot

```

To me this figure is the best results to be discussed regarding REM and LDA

saving plot

```
png("./3_clean_output/figures/PLOT CDAI across time raw.png",  
     width = 8,  
     height = 6,  
     units = "in",  
     res = 120) # opening graphic device
```

CDAI_plot

```
dev.off() # closing graphic device
```

CARRAC histogram

Building large format data table from the CARRAC output

```
# Extracting LDA BARI  
LDA_BARI <- rbind(LDA_BARI_TNF[2,1:4], LDA_BARI_OMA[2,1:4]) # I have  
one estimation per comparison  
LDA_BARI[, LDA := mean(LDA)][, LDA_sup := mean(LDA_sup)][, LDA_inf :=  
mean(LDA_inf)] # averaging  
LDA_BARI <- LDA_BARI[1]
```

```
# REM BARI  
REM_BARI <- rbind(REM_BARI_TNF[2,1:4], REM_BARI_OMA[2,1:4]) # I have  
one estimation per comparison  
REM_BARI[, LDA := mean(LDA)][, LDA_sup := mean(LDA_sup)][, LDA_inf :=  
mean(LDA_inf)] # averaging  
REM_BARI <- REM_BARI[1]
```

```
# IDem for TNFi and OMA  
LDA_TNFi <- LDA_BARI_TNF[3,1:4]  
REM_TNFi <- REM_BARI_TNF[3,1:4]
```

```
LDA_OMA <- LDA_BARI_OMA[3,1:4]  
REM_OMA <- REM_BARI_OMA[3,1:4]
```

```
# Binding together  
LDA <- rbind(LDA_BARI, LDA_TNFi, LDA_OMA)  
setnames(LDA, c("ttt", "LDA", "LDA_sup", "LDA_inf")) #putting right  
labels  
REM <- rbind(REM_BARI, REM_TNFi, REM_OMA)  
setnames(REM, c("ttt", "REM", "REM_sup", "REM_inf"))
```

```
histo_carrac <- cbind(LDA, REM[, -1])
```

plotting

```

1
2
3 carrac_plot <- ggplot(data = histo_carrac, aes(x = ttt, group = ttt))
4 +
5
6   theme_pubclean()+
7
8   geom_errorbar( mapping=aes(x=ttt, ymin=LDA_inf*100,
9 ymax=LDA_sup*100), width=0.2, size=1, color="grey70")+
10   geom_errorbar( mapping=aes(x=ttt, ymin=REM_inf*100,
11 ymax=REM_sup*100), width=0.2, size=1, color="grey50")+
12
13   geom_bar(aes(y = LDA*100), stat = "identity", fill = "grey80", alpha
14 = 0.5)+
15   geom_text(aes(y = LDA*100, label = "LDA"), vjust = 1.5) +
16
17   geom_bar(aes(x = ttt, y = REM*100), stat = "identity", fill =
18 "grey65", alpha = 0.5)+
19   geom_text(aes(x= ttt, y = REM*100, label = "REM"), vjust=1.5) +
20
21   theme(strip.text.y = element_blank(),
22         strip.background = element_blank(),
23         axis.line.x = element_line(size = 0.5),
24         axis.text = element_text(face = "bold", colour = "black"),
25         legend.position="bottom", plot.margin =
26         unit(c(1,3,2,1),"lines"))+
27
28   scale_y_continuous(limits = c(0,82))+
29
30   labs(y = "(% of TC)", x = "Treatment group", title = "A - REM and
31 LDA rates \nby type of treatment \n(CARRAC)")
32
33 carrac_plot
34
35 also Saving CARRAC plot only
36
37 png("./3_clean_output/figures/PLOT BARI 3 CARRAC ONLY.png", width =
38 350, height = 600, res = 100) # opening graphic device
39
40 ggplot(data = histo_carrac, aes(x = ttt, group = ttt)) +
41
42   theme_pubclean()+
43
44   geom_errorbar( mapping=aes(x=ttt, ymin=LDA_inf*100,
45 ymax=LDA_sup*100), width=0.2, size=1, color="grey70")+
46   geom_errorbar( mapping=aes(x=ttt, ymin=REM_inf*100,
47 ymax=REM_sup*100), width=0.2, size=1, color="grey50")+
48
49   geom_bar(aes(y = LDA*100), stat = "identity", fill = "grey80", alpha
50 = 0.5)+
51   geom_text(aes(y = LDA*100, label = "LDA"), vjust = 1.5) +
52
53
54
55
56
57
58
59

```

```
1
2
3   geom_bar(aes(x = ttt, y = REM*100), stat = "identity", fill =
4 "grey65", alpha = 0.5)+
5   geom_text(aes(x= ttt, y = REM*100, label = "REM"), vjust=1.5) +
6
7   theme(strip.text.y = element_blank(),
8         strip.background = element_blank(),
9         axis.line.x = element_line(size = 0.5),
10        axis.text = element_text(face = "bold", colour = "black"),
11        legend.position="bottom", plot.margin =
12 unit(c(1,3,2,1),"lines"))+
13
14   scale_y_continuous(limits = c(0,82))+
15
16   labs(y = "(% of TC)", x = "Treatment group", title = "REM and LDA
17 rates \nby type of treatment \n(CARRAC)")
18
19
20 dev.off() # closing graphic device
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	p2 p2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	p4-5
Objectives	3	State specific objectives, including any prespecified hypotheses	p4-5
Methods			
Study design	4	Present key elements of study design early in the paper	p6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	p6-7
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	p6 p6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	p7 & supp p2
Data sources/measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	supp p2
Bias	9	Describe any efforts to address potential sources of bias	p16
Study size	10	Explain how the study size was arrived at	supp p9
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	p7-8
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	p8-9 supp p 5-6 p8-9 p8-9 p9 & supp p5-6

Continued on next page

Results

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	p10
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	p11-12
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	p13
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	p13-14
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	p14-15 & supp

Discussion

Key results	18	Summarise key results with reference to study objectives	p15
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	p16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	p17
Generalisability	21	Discuss the generalisability (external validity) of the study results	p17

Other information

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	p18
---------	----	---------------------------------------------------------------------------------------------------------------------------------------------------------------	-----

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

Filled in by Benoît GILBERT, 30-01-2023