JAMA Neurology | Original Investigation

Association of Rituximab Treatment With Disability Progression Among Patients With Secondary Progressive Multiple Sclerosis

Yvonne Naegelin, MD; Peter Naegelin; Stefanie von Felten, PhD; Johannes Lorscheider, MD; Judith Sonder, MD; Bernard M. J. Uitdehaag, MD, PhD; Barbara Scotti; Chiara Zecca, MD; Claudio Gobbi, MD; Ludwig Kappos, MD; Tobias Derfuss, MD

IMPORTANCE Therapeutic options for patients with secondary progressive multiple sclerosis (SPMS) are limited.

OBJECTIVE To analyze disability progression in patients with SPMS treated with rituximab compared with matched control patients never treated with rituximab.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study analyzed data obtained from patients with SPMS at 3 multiple sclerosis centers located in Basel and Lugano, Switzerland, and Amsterdam, the Netherlands, from 2004 to 2017. Patients were included for analysis if they had received a diagnosis of SPMS, were treated (57 eligible; 54 included) or never treated (504 eligible; 59 included) with rituximab, and had at least 1 follow-up visit. The variables used for propensity score matching were sex, age, Expanded Disability Status Scale (EDSS) score, and disease duration. Follow-up duration was up to 10 years, with a mean (SD) of 3.5 (2.6) years for rituximab-treated patients and 5.4 (2.4) years for controls in the total cohort and a mean (SD) of 3.5 (2.7) years for rituximab-treated patients and 4.8 (2.2) years for controls in the matched cohort.

EXPOSURES Comparing EDSS score progression in patients with SPMS (treated with rituximab vs not treated with rituximab) using propensity score matching.

MAIN OUTCOMES AND MEASURES The primary end point was progression of EDSS score after baseline, and the secondary end point was time to confirmed disability progression.

RESULTS After 1:1 propensity score matching, 44 matched pairs (88 patients) were included in the analysis. At baseline, patients treated with rituximab had a mean (SD) age of 49.7 (10.0) years, mean (SD) disease duration of 18.2 (9.4) years, and mean (SD) EDSS score of 5.9 (1.4), and 26 (59%) were women, whereas controls had a mean (SD) age of 51.3 (7.4) years, mean (SD) disease duration of 19.4 (8.7) years, and mean (SD) EDSS score of 5.70 (1.29), and 27 (61%) were women. In the covariate-adjusted analysis of the matched set, patients with SPMS who were treated with rituximab had a significantly lower EDSS score during a mean (SD) follow-up of 3.5 (2.7) years (mean difference, −0.52; 95% CI, −0.79 to −0.26; $P < .001$). Time to confirmed disability progression was significantly delayed in the rituximab-treated group (hazard ratio, 0.49; 95% CI, 0.26-0.93; $P = .03$).

CONCLUSIONS AND RELEVANCE In this study, patients with SPMS treated with rituximab had a significantly lower EDSS score for up to 10 years of follow-up and a significantly delayed confirmed progression compared with matched controls, suggesting that B-cell depletion by rituximab may be therapeutically beneficial in these patients. A prospective randomized clinical trial with a better level of evidence is needed to confirm the efficacy of rituximab in such patients.

JAMA Neurol. 2019;76(3):274-281. doi:10.1001/jamaneurol.2018.4239
Published online January 7, 2019.

Supplemental content

Author Affiliations: Neurologic Clinic and Policlinic, University Hospital and University of Basel, Basel, Switzerland (Y. Naegelin, P. Naegelin, Lorscheider, Kappos, Derfuss); Clinical Trial Unit, Department of Clinical Research, University of Basel, University Hospital Basel, Basel, Switzerland (von Felten); MS Center Amsterdam, Department of Neurology, VU University Medical Center, Amsterdam, the Netherlands (Sonder, Uitdehaag); Multiple Sclerosis Center, Neurocenter of Southern Switzerland, Ospedale Regionale di Lugano, Lugano, Switzerland (Scotti, Zecca, Gobbi).

Corresponding Author: Yvonne Naegelin, MD, Neurologic Clinic and Policlinic, University Hospital and University of Basel, Petersgraben 4, 4031 Basel, Switzerland (yvonne.naegelin@usb.ch).

Multiple sclerosis (MS) is a chronic inflammatory demyelinating disease of the central nervous system (CNS).[1] The majority of patients present with a relapsing-remitting (RRMS) course of disease, followed by a progressive phase termed secondary progressive MS (SPMS).[2] Current disease-modifying treatments target the inflammatory pathology and only indirectly the neurodegenerative pathology of the disease, and their therapeutic effects in (secondary) progressive MS have been very limited.

Data from animal models, human neuropathologic studies, and clinical trials suggest a prominent role for B cells in the pathogenesis of MS, both in RRMS and SPMS. The progressive phase is characterized by a compartmentalized inflammatory process that persists beyond a relatively intact blood-brain barrier. B-cell follicle-like structures are found in the meninges as a correlate of this CNS restricted inflammation.[1,3] Potential drugs for progressive forms of MS should, therefore, be able to pass the blood-brain barrier and should be able to target proinflammatory mediators and mechanisms directly in the brain or should influence the immune axis between the peripheral immune system and the CNS.[3] Recent studies postulate that there is a B-cell exchange across the blood-brain barrier. Rituximab, a monoclonal CD20 antibody, might affect the B-cell population within the CNS through depletion of the peripheral B-cell compartment.[4,5] Rituximab is detectable in low concentrations within cerebrospinal fluid after intravenous administration, opening the possibility of a direct effect on CNS resident B cells.[6]

In a phase 2 trial, rituximab reduced clinical activity and inflammatory brain lesions in RRMS.[7] Another trial tested rituximab in primary progressive multiple sclerosis (PPMS).[8] A post hoc analysis showed that rituximab affected disability progression in patients of younger age and those with contrast-enhancing lesions.

More recently the humanized monoclonal anti-CD20 antibody ocrelizumab significantly reduced the percentage of patients with confirmed disability progression (CDP) compared with interferon beta-1a in RRMS and significantly reduced CDP compared with placebo in PPMS.[9,10] In 2014, the fully humanized monoclonal anti-CD20 antibody ofatumumab also demonstrated efficacy in a phase 2 clinical study of patients with RRMS.[11]

In a recently published Swedish study, the authors concluded that treatment with rituximab was safe after a follow-up of up to 2 years.[12] Owing to the limited treatment options for SPMS and the extrapolation of results in RRMS and PPMS, rituximab was used off-label for the treatment of SPMS. The present retrospective cohort study uses propensity score matching to compare disease progression between patients who were treated with rituximab and patients who had never been treated with rituximab.

Key Points

Question Does disability progression among patients with secondary progressive multiple sclerosis treated with the monoclonal anti-CD20 antibody rituximab differ from that among such patients never treated with rituximab?

Findings In this cohort study of 88 propensity score–matched patients, those treated with rituximab had a significantly lower Expanded Disability Status Scale score for up to 10 years of follow-up and significantly delayed confirmed progression compared with matched controls. No associations between confirmed progression and individual patient baseline characteristics were identified.

Meaning Therapeutic options for patients with secondary progressive multiple sclerosis are limited; however, these findings suggest that B-cell–depleting therapy may be beneficial.

Methods

Patients with SPMS, defined according to established criteria,[13] were eligible for this retrospective study if they had been treated with rituximab off-label at the MS Centers in Basel or Lugano, Switzerland; had received at least 1 dose of rituximab; had had at least 1 clinical follow-up visit; and had provided informed consent (see **Figure 1** for the study flowchart). Patients were treated with rituximab based on the decision of the treating neurologist that the disease was progressive and that no other treatment was available. The study was approved by the local ethics committees of Lugano (*comitato etico cantonale*, Bellinzona) and Basel (*Ethikkommission Nordwestschweiz und Zentralschweiz* [EKNZ]) in Switzerland. Informed consent was obtained in written (Lugano) or verbal (by telephone interview, Basel) format.

In total, 54 rituximab-treated patients with SPMS were compared with 59 patients with SPMS who had never been treated with rituximab (control group) and were part of an observational cohort study conducted at the MS Center Basel, University of Basel, and the MS Center Amsterdam (at University Medical Center Amsterdam, the Netherlands). That study, including the informed consent procedure, was approved by the local ethics committees of Basel (EKNZ), and Amsterdam (Medical Ethical Committee VUmc), and all patients provided written informed consent.

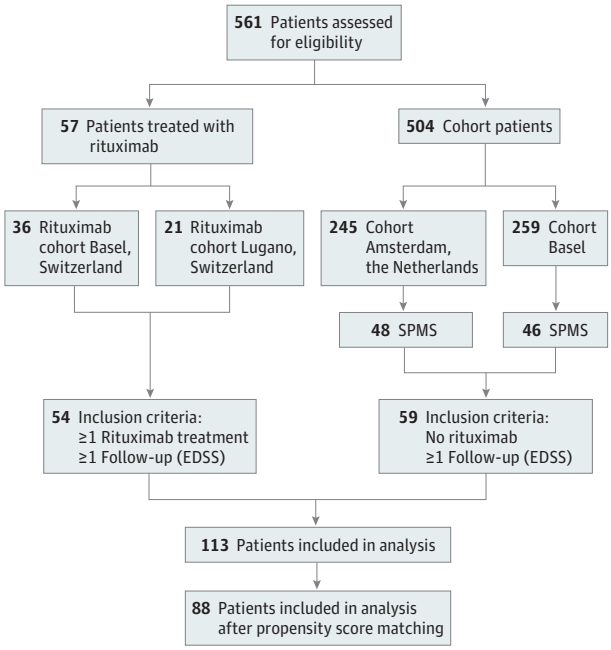
In a second step, the rituximab-treated and the control groups were matched 1:1 using propensity scores, resulting in 44 matched patient pairs. The matching variables were sex, age, Expanded Disability Status Scale (EDSS) score, and disease duration at baseline.

For all patients, (at least) yearly standardized clinical assessments were performed, including a full neurologic examination (Neurostatus-EDSS). In addition, routine magnetic resonance imaging (MRI) findings for new T2 lesions and gadolinium-enhancing lesions were analyzed at baseline. Because MRIs at baseline were available for only 33 patients (approximately 60% of the total cohort), MRI findings could not be used for the matching process. Magnetic resonance imaging during follow-up was performed in only a minority of patients, and the images were therefore not analyzed for the present study.

Main Outcome Measures

The primary objective of the present study was to compare disease progression as assessed by the EDSS score after baseline

Figure 1. Study Flowchart



EDSS indicates Expanded Disability Status Scale; SPMS, secondary progressive multiple sclerosis.

in patients with SPMS who were treated with rituximab vs those who were not treated with rituximab. The EDSS score was determined annually for up to 10 years after baseline (**Table 1**). For patients treated with rituximab, EDSS scores acquired during treatment and for up to 1 year after the end of treatment were analyzed.

One secondary objective was to analyze the time to CDP in rituximab-treated patients with SPMS vs matched control patients. Confirmed progression was defined as an increase in the EDSS score 12 or more months after baseline, which was confirmed by a second examination conducted 12 months later. The increase had to be at least 1.5 steps for an EDSS score of 0, 1 step for scores between 1 and 5, and 0.5 steps for scores of 5.5 or greater.

The other secondary objective was to compare baseline characteristics of patients in the rituximab-treated group with confirmed progression with those of patients without confirmed progression in the same group.

Statistical Analysis

An initial comparison of baseline characteristics (age, sex, EDSS score, and disease duration) between the 2 groups showed that patients treated with rituximab were significantly younger and had a higher grade of disability (Table 1). To improve the balance of baseline characteristics between the 2 groups, propensity score matching as implemented in the R package nonrandom software was used to create a 1:1 matched data set of 44 rituximab-treated patients and 44 controls.[14] Density plots of the distribution of propensity scores per cohort before and after matching are shown in eFigure 1 in the Supplement. Standardized differences between cohorts before and after propensity score matching were calculated (eFigure 2 and eTable 1 in the Supplement).

Statistical analysis was performed using the total cohort as well as using the matched cohort. Although matching improved the balance, considerable differences in baseline characteristics remained between the matched groups. Thus, all baseline characteristics were included as covariates in the analysis of the total cohort as well as in the analysis of the matched cohort. The primary end point, EDSS score after baseline, was analyzed using a linear mixed-effects model to estimate effect sizes together with 95% CIs. To account for multiple measurements per patient, a random intercept was added for each patient. As explanatory variables, the model included age, sex, disease duration, baseline EDSS score, treatment (rituximab vs control), time after baseline, and the interaction between treatment and time after baseline. All continuous explanatory variables (age, disease duration, and time after baseline) and baseline EDSS scores were centered by subtracting the respective median value for better interpretation of model intercepts.

We used a Bayesian approach to graphically display the fitted values together with their credible intervals (estimated by the corresponding model) for the primary end point. Posterior distributions of the fitted values were calculated using the sim function in the arm package of R.[15] The fitted values and credible intervals are displayed for a female "model patient" (the present data set included more women than men) with median values of the baseline EDSS score, baseline age, and disease duration (to represent the data set as accurately as possible).

Time to confirmed progression was analyzed as a secondary end point. Because there were 9 patients with 2 confirmed progression events (all in the control group), we used recurrent event analysis by marginal means and rates models, allowing the accounting of multiple events per patient. Age, sex, disease duration, baseline EDSS score, and treatment were used as explanatory variables. A "cluster" term in the model was used to compute a robust variance for the model, accounting for nonindependent events from the same patient. Moreover, we compared rituximab-treated patients with or without confirmed progression regarding baseline characteristics and rituximab treatments. Because the number of rituximab-treated patients with confirmed progression was small (n = 12), only bivariate associations between confirmed progression and individual patient characteristics were assessed. Associations between confirmed progression and categorical variables (eg, sex) were assessed using the Fisher exact test. Associations between confirmed progression and continuous variables were assessed using a Wilcoxon rank sum test.

All statistical analyses were performed with the statistical software environment R, version 3.4.4 (R Core Team 2018).[16] A 2-sided $P < .05$ was considered statistically significant.

Results

Table 1 shows the baseline characteristics for 54 rituximab-treated patients and 59 control patients who were not treated

Table 1. Baseline Characteristics and Follow-up Duration for the Total Cohort

Characteristic[a]	Group Receiving Off-label Rituximab Treatment	Control Group	*P* Value[b]
Before propensity score matching			
No. of patients	54	59	
Age, y			
Mean (SD)	49.0 (9.6)	53.5 (8.0)	.008
Median (range)	49.0 (23.0-71.0)	54.0 (37.0-70.0)	
Female, No. (%)	32 (59)	33 (56)	.85
Disease duration, y			
Mean (SD)	18.6 (9.3)	18.8 (9.0)	.97
Median (range)	18.0 (3.0-40.0)	18.0 (3.0-45.0)	
EDSS BL			
Mean (SD)	6.02 (1.32)	5.21 (1.47)	.002
Median (range)	6.25 (2.50-8.50)	5.50 (2.00-8.00)	
Follow-up, mo			
Mean (SD)	42.3 (31.4)	64.3 (29.4)	
Median (range)	33.9 (6.6-111.4)	59.9 (12.2-123.5)	
Follow-up, y			<.001
Mean (SD)	3.5 (2.6)	5.4 (2.4)	
Median (range)	2.8 (0.6-9.3)	5.0 (1.0-10.3)	
After propensity score matching			
No. of patients	44	44	
Age, y			
Mean (SD)	49.7 (10.0)	51.3 (7.4)	.36
Median (range)	50.0 (23.0-71.0)	50.5 (37.0-65.0)	
Female, No. (%)	26 (59)	27 (61)	>.99
Disease duration, y			
Mean (SD)	18.2 (9.4)	19.4 (8.7)	.45
Median (range)	17.0 (3.0-40.0)	19.5 (3.0-35.0)	
EDSS BL			
Mean (SD)	5.93 (1.40)	5.70 (1.29)	.28
Median (range)	6.00 (2.50-8.50)	6.00 (2.50-8.00)	
Follow-up, mo			.
Mean (SD)	41.8 (32.2)	57.7 (26.5)	
Median (range)	29.8 (6.6-111.4)	54.5 (12.2-112.3)	
Follow-up, y			002
Mean (SD)	3.5 (2.7)	4.8 (2.2)	
Median (range)	2.5 (0.6-9.3)	4.5 (1.0-9.4)	

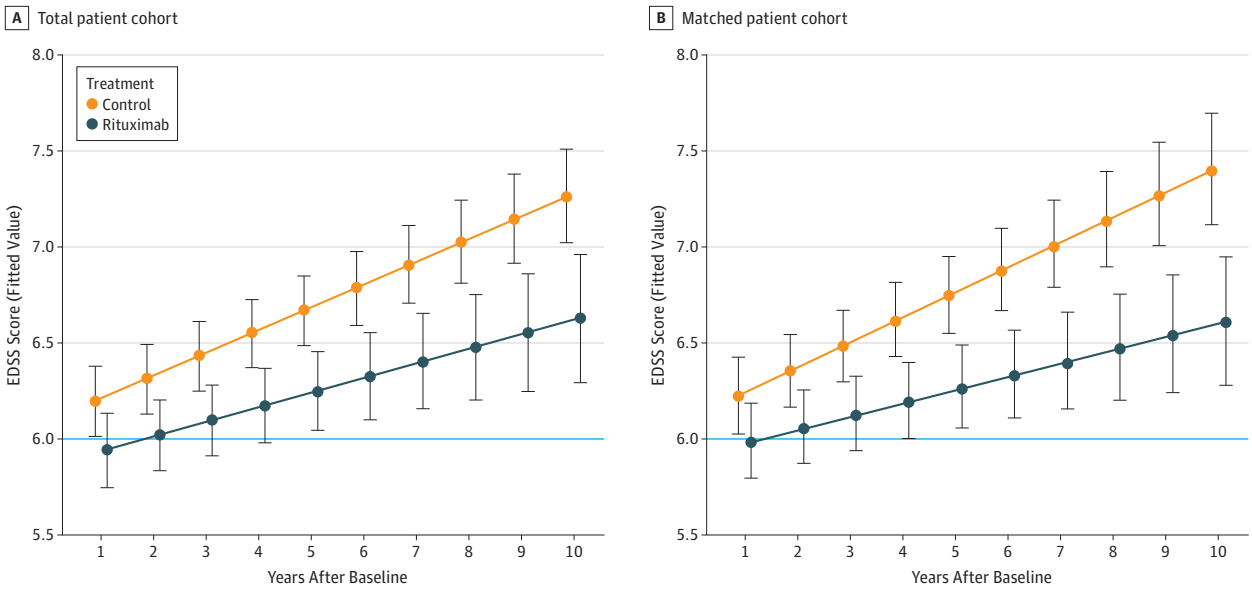
Abbreviation: EDSS BL, Expanded Disability Status Scale score after baseline.

[a] Number and percentage of patients with measurements are given for categorical variables; mean (SD) and median (range) are given for continuous and ordinal variables.

[b] Derived from Fisher exact tests (categorical) and from Wilcoxon rank sum tests (Mann-Whitney tests, continuous). For standardized differences between groups before and after propensity score matching, see eTable 1 in the Supplement.

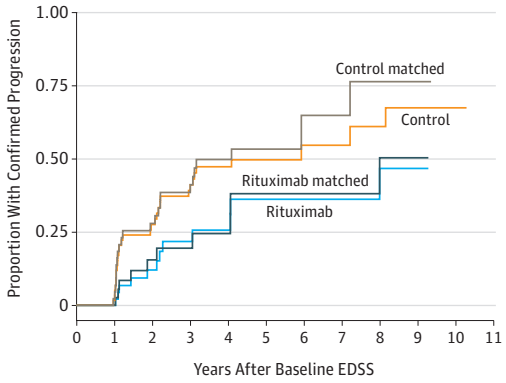
with rituximab before and after matching. Before matching, patients from the rituximab group were significantly younger and had a higher grade of disability. After matching, there were no significant differences between the groups; however, *P* values should be compared with caution because group sizes were reduced. At baseline after matching, patients treated with rituximab had a mean (SD) age of 49.7 (10.0) years, mean (SD) disease duration of 18.2 (9.4) years, and mean (SD) EDSS score of 5.93 (1.40), and 26 (59%) were women, whereas controls had a mean (SD) age of 51.3 (7.4) years, mean (SD) disease duration of 19.4 (8.7) years, and a mean (SD) EDSS score of 5.70 (1.29), and 27 (61%) were women. Standardized differences before and after matching showed improved balance for the covariates EDSS score and age after matching (eFigure 2 and eTable 1 in the Supplement). Analysis of MRI findings at baseline, which were not available for all patients (total rituximab group, 33 [61%]; matched rituximab group, 27 [61%]; total control group, 35 [59%]; matched control group, 23 [52%]), revealed that 7 patients (26%) in the matched rituximab group and 0 patients in the matched control group had gadolinium-enhancing lesions. Baseline MRIs for the matched rituximab group were performed a mean (SD) of 84.9 (57.8) days before baseline (median, 77 days; range, 0-198 days), whereas baseline MRIs in the control group were performed at baseline. Of 8 patients treated with rituximab (rituximab total group) with an MRI finding of active disease at baseline, only 1 patient developed CDP during the study period. In the rituximab group (matched cohort), 18 patients (41%) had received no immunomodulatory treatment during the year before baseline, whereas in the control group, a slightly higher proportion (21 patients; 48%;) was untreated (for a detailed treatment summary see eTable 2 in the Supplement).

Figure 2. Progression of Expanded Disability Status Scale (EDSS) Score Over Time for Both Cohorts (Total and Matched)



Fitted values of the EDSS scores together with Bayesian 95% credible intervals 1 to 10 years after baseline for the total cohort (A) and the matched cohort (B). The fitted values represent a female model patient of median age at baseline (A, 51 years; B, 50 years), median disease duration (A, 18 years; B, 18.5 years), and with median baseline EDSS score (A, 6.0; B, 6.0). Median baseline EDSS score is shown as a horizontal blue line. The mean treatment difference (during the follow-up period) between rituximab and control of −0.45 (95% CI, −0.71 to −0.20), estimated by the statistical model on the total set, can be read from panel A at 5.5 years after baseline. Likewise, the mean treatment difference of −0.52 (95% CI, −0.79 to −0.26) for the matched cohort can be read from panel B.

Figure 3. Kaplan-Meier Curve for Time to First Confirmed Progression



No. at risk

Rituximab	54	47	30	21	14	10	8	7	6	2	0
Rituximab matched	44	38	21	17	11	8	6	6	5	1	0
Control	59	56	40	30	23	14	9	7	6	5	2
Control matched	44	42	29	22	15	7	3	3	2	2	0

EDSS indicates Expanded Disability Status Scale.

The analysis of the primary end point showed that rituximab treatment significantly reduced progression of disability as measured by yearly EDSS scores for a follow-up of up to 10 years (mean [SD] follow-up, 3.5 [2.6] years) in the total patient cohort (estimate, −0.45; 95% CI, −0.71 to −0.20; $P < .001$) as well as in the matched cohort (mean [SD] follow-up, 3.5 [2.7] years; estimate, −0.52; 95% CI, −0.79 to −0.26; $P < .001$). **Figure 2** shows progression of EDSS scores over time for both cohorts (total and matched) as estimated by the statistical models.

With respect to the secondary end point, time to confirmed progression was significantly longer in the rituximab-treated group than in the control group for the total cohort (hazard ratio, 0.48; 95% CI, 0.26-0.91; $P = .02$) and the matched cohort (hazard ratio, 0.49; 95% CI, 0.26-0.93; $P = .03$). **Figure 3** shows the Kaplan-Meier plots for time to first confirmed progression.

In the rituximab-treated group, baseline characteristics as well as number and dose of rituximab treatments for 12 patients with or 42 without progression were compared to identify potential predictors for treatment response. **Table 2** shows the categorical (sex and reason for rituximab) and continuous (age, EDSS score at baseline, disease duration, number of rituximab cycles, total cumulative dose of administered rituximab, duration of treatment, and treatment intensity) patient characteristics for rituximab-treated patients with or without confirmed progression. Patients with confirmed progression had higher cumulative doses of rituximab and more treatment cycles but for a longer period of time; thus, the treatment intensity (defined as the total amount of rituximab administered in milligrams divided by the duration of treatment in months) remained similar compared with patients without confirmed progression. In summary, no associations between confirmed progression and individual patient baseline characteristics could be identified.

At the cutoff time for this analysis, 29 of 54 patients (54%) were still being treated with rituximab with no complica-

Table 2. Characteristics of Patients Administered Rituximab With or Without Confirmed Progression

Characteristic[a]	Rituximab With Confirmed Progression	Rituximab Without Confirmed Progression	*P* Value[b]
Categorical			
No. of patients	12	42	
Female, No. (%)	7 (58)	25 (60)	>.99
Reason for rituximab, No. (%)			
Progression	11 (92)	26 (62)	.16
Relapses	1 (8)	3 (7)	
Progression and relapses	0	3 (7)	
Other	0	0 (24)	
Relapses under treatment, No. (%)	1 (8)	5 (12)	>.99
Continuous			
No. of patients	12	42	
Age, y			
Mean (SD)	49.7 (10.6)	48.8 (9.4)	.95
Median (range)	48.0 (33.0-71.0)	49.0 (23.0-70.0)	
Disease duration, y			
Mean (SD)	19.0 (10.6)	18.6 (9.0)	.83
Median (range)	19.5 (3.0-40.0)	18.0 (3.0-39.0)	
Rituximab, No. of cycles			
Mean (SD)	6.00 (2.09)	4.29 (2.99)	.04
Median (range)	6.50 (3.0-9.00)	3.50 (1.00-11.00)	
Rituximab, cumulative dose, mg			
Mean (SD)	9.54 (3.86)	6.00 (3.85)	.007
Median (range)	8.50 (5.00-15.00)	5.00 (1.00-16.00)	
Rituximab, treatment duration, mo			
Mean (SD)	60.9 (27.2)	42.1 (30.7)	.03
Median (range)	49.5 (32.0-110.0)	34.5 (6.0-107.0)	
Rituximab, treatment intensity, mg/mo[c]			
Mean (SD)	161.03 (32.10)	160.32 (50.85)	.88
Median (range)	156.05 (131.82-250.00)	160.26 (67.31-333.33)	
Ordinal			
EDSS BL			
Mean (SD)	5.71 (1.41)	6.11 (1.30)	.35
Median (range)	6.00 (3.00-7.50)	6.50 (2.50-8.50)	

Abbreviation: EDSS BL, Expanded Disability Status Scale score after baseline.

[a] Number and percentage of patients with measurements are given for categorical variables; mean (SD) and median (range) are given for continuous variables.

[b] Derived from Fisher exact tests (categorical) and from Wilcoxon rank sum tests (Mann-Whitney tests, continuous).

[c] Defined as the total dose in milligrams of rituximab administered divided by the duration of treatment in months.

tions. In 5 cases (9%), complications were reported: 1 patient had leukocytoclastic vasculitis in both legs (confirmed by biopsy) after the first infusion (as reported previously[17]), 1 patient had a segmental herpes zoster infection, and 3 had 1 or more pneumonia events or urinary tract infections.

Two patients died during the follow-up period: 1 died of a spontaneous intracerebral hemorrhage 3 years after stopping rituximab treatment (rituximab therapy was stopped because the disease was stabilized), and 1 died of pneumonia 4 years after rituximab treatment had been stopped.

Discussion

To date, no prospective clinical trial has been performed to evaluate the efficacy of B-cell depletion in SPMS. The main goal of the present study was to compare disease progression in patients with SPMS treated with rituximab with that in matched control patients in a comprehensive real-world cohort. In the rituximab group, we observed a significant reduction of EDSS score progression after baseline, the study's primary end point. This reduction in disability progression as assessed annually by the EDSS score for up to 10 years was significant in both the total (mean follow-up, 3.5 years) and the propensity score–matched (mean follow-up, 3.5 years) cohorts. The time to confirmed progression was also significantly delayed in the rituximab-treated group compared with the control group for the total and matched cohorts. These findings suggest that rituximab administration may be associated with a beneficial therapeutic effect in patients with SPMS.

Approximately 75% of untreated and 50% of treated individuals in our cohorts developed clinically significant confirmed progression for the 10-year period (Figure 3). This is in line with other cohorts in which the rates of progression during a 3- and a 10-year follow-up were 15% and 75%, respectively.[18-20]

The baseline characteristics of our patient cohort closely resembled those of patients included in the recently completed clinical trial examining the efficacy and safety of siponimod in SPMS (EXPAND).[21] The EXPAND study showed a significant reduction of disability progression in the treatment group compared with that in the placebo group. The percentages of patients with a 6-month CDP were for placebo and siponimod 28% and 20% (year 2) and 30% and 23% (year 3), respectively. This is comparable to a CDP in 25% of matched controls and in 12.5% of rituximab-treated patients at year 2 and in 37.5% of matched controls and in 25% of rituximab-treated patients at year 3 in our cohort. Another study looking at the efficacy of natalizumab for reducing disability progression in participants with SPMS (ASCEND)[22] did not meet the primary end point of reducing CDP. In a retrospective, propensity score–matched comparison of patients with SPMS from the MSBase registry who did not show an effect of immunomodulatory treatment (mainly interferon beta and glatiramer acetate, no patients treated with rituximab were included in the matched data set analysis) on disease progression, the rate of 6-month confirmed disease progression after 3 years was 23% (J.L., written communication, September 28, 2018).[23] These studies, along with our small retrospective analysis, suggest that only a few anti-inflammatory treatments may be associated with a beneficial outcome in patients with SPMS. All analyzed compounds described above influence B-cell biology albeit by different mechanisms of action. The differential response to these treatments may provide clues to understanding which parts of the B-cell response are pathogenic in SPMS and which patients might benefit from such treatments.

It is crucial to identify patient characteristics that indicate a higher chance of treatment response. To identify potential prognostic markers for therapeutic response, baseline characteristics within the rituximab-treated group, comparing patients with or without confirmed progression, were tested. There was no significant difference in any tested category. However, MRI scans were available in only 60% of patients; therefore, these results should be interpreted with caution. Of 8 patients treated with rituximab (rituximab total group) with an MRI finding of active disease at baseline, only 1 patient developed CDP during the study period, whereas 12 of 54 patients in the rituximab total group showed CDP during the present study. This might indicate that patients with active disease respond better to this kind of treatment. In general, gadolinium-enhancing lesions at baseline are a risk factor for CDP in PPMS.[10,24] The imbalance of increased MRI lesion activity in the rituximab group of our cohort might therefore favor the control group and underestimate the treatment effects associated with administration of rituximab. Whether MRI lesion activity is also associated with treatment response must be evaluated in larger cohorts with standardized MRI protocols.

There were no major safety concerns during the treatment period, but complications were documented in 5 cases (9.%), mainly related to infections.

Limitations

This study has some limitations. First, it was a retrospective analysis of clinical data and not a prospective, controlled, randomized clinical trial. The present study design has a higher risk of confounding the treatment effect associated with rituximab administration with other patient documented or unknown characteristics and, as a consequence, a biased estimate of the association. In fact, review of the baseline characteristics indicated that patients in the rituximab group tended to be younger, had a higher EDSS score, and showed more MRI lesion activity. It is conceivable that these patients were more likely to be treated with rituximab.

In addition, that a higher proportion of the control patients had not been treated with disease-modifying treatments in the year before baseline (49% and 48% in the total and matched groups, respectively, compared with 39% and 41% of the rituximab-treated patients) could indicate a bias toward a less aggressive disease course in the control group that may contribute to underestimating the effects associated with administration of rituximab. To mitigate confounding effects of known disease characteristics, we used propensity score–based matching in combination with covariate adjustment in the statistical models. As shown by the standardized differences before and after matching, the differences between the groups were smaller after propensity score matching. Furthermore, it was reassuring that the significant association observed in the propensity score–matched comparison was also significant in the total group analysis.

Conclusions

Patients with SPMS who were treated with rituximab had a significantly lower EDSS score up to 10 years after the initial assessment and significantly delayed progression compared with patients with SPMS who had never been treated with rituximab. A prospective randomized clinical trial is needed to show efficacy in this patient group with a higher level of evidence.

ARTICLE INFORMATION

Accepted for Publication: November 2, 2018.

Published Online: January 7, 2019. doi:10.1001/jamaneurol.2018.4239

Author Contributions: Dr Y. Naegelin had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Y. Naegelin, Kappos, Derfuss.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Y. Naegelin, P. Naegelin.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Y. Naegelin, P. Naegelin, von Felten.
Administrative, technical, or material support: Y. Naegelin, P. Naegelin.
Supervision: Y. Naegelin, Kappos, Derfuss.

Conflict of Interest Disclosures: Dr Lorscheider reported receiving research support from Biogen Inc and serving as a paid member of the advisory board for F. Hoffmann-La Roche Ltd. Dr Uitdehaag reported receiving research support and personal compensation for consulting from Biogen Inc, Sanofi Genzyme, Merck Serono, Novartis, Roche and TEVA. Dr Zecca reported receiving financial support from Almirall, Biogen Inc, Bayer Schering Pharma, Celgene, Sanofi Genzyme, Merck Serono, Novartis, Teva, and F. Hoffmann-La Roche Ltd outside the submitted work. Dr Gobbi reported receiving financial support from Almirall, Biogen Inc, Bayer Schering Pharma, Celgene, Sanofi Genzyme, Merck Serono, Novartis, Teva, and

F. Hoffmann-La Roche Ltd outside the submitted work. Dr Kappos reported receiving research support from the Swiss Multiple Sclerosis Society, Swiss National Research Foundation, European Union, and Roche Research Foundation. Dr Kappos reported that his employer, the University Hospital of Basel, has received and used exclusively for research support compensation fees for his participation on steering committees or as an advisory board member or consultant of Actelion, Alkermes, Allergan, Almirall, Bayer Health Care, Biogen Inc, Celgene, CSL Behring, Excemed, df-mp, GeNeuro SA, Sanofi Genzyme, Merck Serono, Mitsubishi Tanabe Pharma, Novartis, Pfizer, F. Hoffmann-La Roche Ltd, Sanofi-Aventis, Santhera Pharmaceuticals, Teva, UCB, Vianex SA, and Xenoport, and also for license fees for Neurostatus products. Dr Derfuss reported serving on scientific advisory boards for Novartis, Merck Serono, Biogen Inc, Sanofi Genzyme, GeNeuro, MedDay, Mitsubishi Tanabe Pharma, F. Hoffmann-La Roche Ltd, and Bayer Schering Pharma; receiving funding for travel or speaker honoraria from Biogen Inc, Sanofi Genzyme, Novartis, Merck Serono, F. Hoffmann-La Roche Ltd, and Bayer Schering Pharma; and receiving research support from Biogen Inc, Novartis, the European Union, the Swiss National Foundation, and the Swiss Multiple Sclerosis Society. No other disclosures were reported.

Additional Contributions: We thank all patients for their participation in this study.

REFERENCES

1. Lassmann H, Brück W, Lucchinetti CF. The immunopathology of multiple sclerosis: an overview. *Brain Pathol*. 2007;17(2):210-218. doi:10.1111/j.1750-3639.2007.00064.x

2. Confavreux C, Vukusic S, Moreau T, Adeleine P. Relapses and progression of disability in multiple sclerosis. *N Engl J Med*. 2000;343(20):1430-1438. doi:10.1056/NEJM200011163432001

3. Mahad DH, Trapp BD, Lassmann H. Pathological mechanisms in progressive multiple sclerosis. *Lancet Neurol*. 2015;14(2):183-193. doi:10.1016/S1474-4422(14)70256-X

4. von Büdingen HC, Kuo TC, Sirota M, et al. B cell exchange across the blood-brain barrier in multiple sclerosis. *J Clin Invest*. 2012;122(12):4533-4543. doi:10.1172/JCI63842

5. Palanichamy A, Apeltsin L, Kuo TC, et al. Immunoglobulin class-switched B cells form an active immune axis between CNS and periphery in multiple sclerosis. *Sci Transl Med*. 2014;6(248):248ra106. doi:10.1126/scitranslmed.3008930

6. Rubenstein JL, Combs D, Rosenberg J, et al. Rituximab therapy for CNS lymphomas: targeting the leptomeningeal compartment. *Blood*. 2003;101(2):466-468. doi:10.1182/blood-2002-06-1636

7. Hauser SL, Waubant E, Arnold DL, et al; HERMES Trial Group. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. *N Engl J Med*. 2008;358(7):676-688. doi:10.1056/NEJMoa0706383

8. Hawker K, O'Connor P, Freedman MS, et al; OLYMPUS trial group. Rituximab in patients with primary progressive multiple sclerosis: results of a randomized double-blind placebo-controlled multicenter trial. *Ann Neurol*. 2009;66(4):460-471. doi:10.1002/ana.21867

9. Hauser SL, Bar-Or A, Comi G, et al; OPERA I and OPERA II Clinical Investigators. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. *N Engl J Med*. 2017;376(3):221-234. doi:10.1056/NEJMoa1601277

10. Montalban X, Hauser SL, Kappos L, et al; ORATORIO Clinical Investigators. Ocrelizumab versus placebo in primary progressive multiple sclerosis. *N Engl J Med*. 2017;376(3):209-220. doi:10.1056/NEJMoa1606468

11. Sorensen PS, Lisby S, Grove R, et al. Safety and efficacy of ofatumumab in relapsing-remitting multiple sclerosis: a phase 2 study. *Neurology*. 2014;82(7):573-581. doi:10.1212/WNL.0000000000000125

12. Salzer J, Svenningsson R, Alping P, et al. Rituximab in multiple sclerosis: a retrospective observational study on safety and efficacy. *Neurology*. 2016;87(20):2074-2081. doi:10.1212/WNL.0000000000003331

13. Lublin FD, Reingold SC; National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology*. 1996;46(4):907-911. doi:10.1212/WNL.46.4.907

14. Stampf S. Package "nonrandom": stratification and matching by the propensity score. https://rdrr.io/cran/nonrandom/. Published February 20, 2015. Accessed November 21, 2018.

15. Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press; 2007.

16. R Core Team. R: a language and environment for statistical computing. The R project for statistical computing. Vienna, Austria. https://www.R-project.org/. Published 2018. Accessed November 27, 2018.

17. Kandula P, Kouides PA. Rituximab-induced leukocytoclastic vasculitis: a case report. *Arch Dermatol*. 2006;142(2):246-247. doi:10.1001/archderm.142.2.246

18. Bove R, Chitnis T, Cree BA, et al; SUMMIT Consortium. SUMMIT (Serially Unified Multicenter Multiple Sclerosis Investigation): creating a repository of deeply phenotyped contemporary multiple sclerosis cohorts. *Mult Scler*. 2018;24(11):1485-1498. doi:10.1177/1352458517726657

19. Healy BC, Engler D, Gholipour T, Weiner H, Bakshi R, Chitnis T. Accounting for disease modifying therapy in models of clinical progression in multiple sclerosis. *J Neurol Sci*. 2011;303(1-2):109-113. doi:10.1016/j.jns.2010.12.024

20. Cree BA, Gourraud PA, Oksenberg JR, et al; University of California, San Francisco MS-EPIC Team. Long-term evolution of multiple sclerosis disability in the treatment era. *Ann Neurol*. 2016;80(4):499-510. doi:10.1002/ana.24747

21. Kappos L, Bar-Or A, Cree BAC, et al; EXPAND Clinical Investigators. Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study. *Lancet*. 2018;391(10127):1263-1273. doi:10.1016/S0140-6736(18)30475-6

22. US National Library of Medicine. A clinical study of the efficacy of natalizumab on reducing disability progression in participants with secondary progressive multiple sclerosis (ASCEND in SPMS). https://clinicaltrials.gov/ct2/show/NCT01416181. Accessed November 21, 2018.

23. Lorscheider J, Jokubaitis VG, Spelman T, et al; MSBase Study Group. Anti-inflammatory disease-modifying treatment and short-term disability progression in SPMS. *Neurology*. 2017;89(10):1050-1059. doi:10.1212/WNL.0000000000004330

24. Stevenson VL, Ingle GT, Miller DH, Thompson AJ. Magnetic resonance imaging predictors of disability in primary progressive multiple sclerosis: a 5-year study. *Mult Scler*. 2004;10(4):398-401. doi:10.1191/1352458504ms1055oa

Supplementary Online Content

Naegelin Y, Naegelin P, von Felten S, et al. Association of rituximab treatment with disability progression among patients with secondary progressive multiple sclerosis. *JAMA Neurol*. Published online January 7, 2019. doi:10.1001/jamaneurol.2018.4239

This supplementary material has been provided by the authors to give readers additional information about their work.

eTable 1. Standardized Differences

Before Matching			
Variable	**Std. Diff**	**Std. Diff. Low**	**Std. Diff. Up**
Age	0.51	0.13	0.88
EDSS	0.58	0.20	0.95
Disease Duration	0.01	-0.35	0.38
Follow-up time	0.72	0.34	1.10
Gender	0.07	-0.30	0.44
After Matching			
Variable	**Std. Diff**	**Std. Diff. Low**	**Std. Diff. Up**
Age	0.19	-0.23	0.60
EDSS	0.17	-0.25	0.59
Disease Duration	0.13	-0.29	0.55
Follow-up time	0.54	0.11	0.96
Gender	0.05	-0.37	0.46

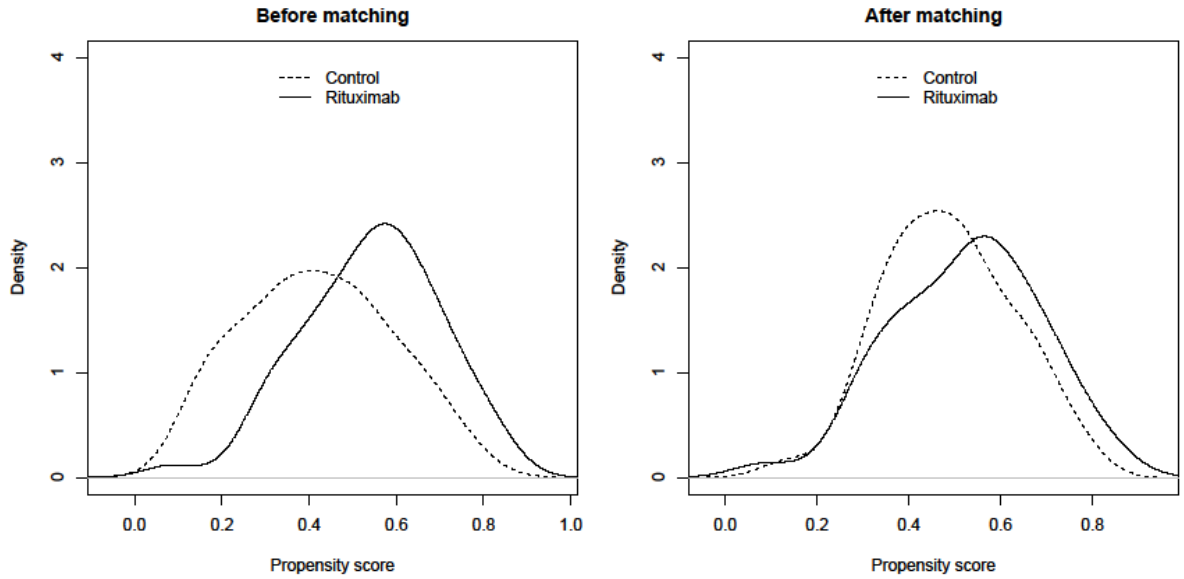
eTable 1: Standardized differences for baseline variables (used for matching) and follow-up time (in addition) before and after matching

eTable 2. Treatment

Cohort	Pre-BL (1 year before BL)			BL and Follow-up				
Rituximab matched	**Medication**	**n= 44**	**%**	**Medication BL n=44**				
	None	18	40.9	Rituximab				
	INF-β 1a i.m.	2	4.5					
	INF-β 1b	4	9.1					
	GA	4	9.1					
	Fingolimod	2	4.5					
	Methotrexat	0	0					
	Mitoxantrone	6	13.7					
	Mycophenolat-Mophetil	1	2.3					
	Natalizumab	4	9.1					
	INF-β 1a s.c.	2	4.5					
	Rituximab	0	0					
	Teriflunomide	1	2.3					
Rituximab all	**Medication**	**n= 54**	**%**	**Medication BL n=54**				
	None	21	38.9	Rituximab				
	INF-β 1a i.m.	2	3.7					
	INF-β 1b	4	7.4					
	GA	6	11.1					
	Fingolimod	4	7.4					
	Methotrexat	1	1.9					
	Mitoxantrone	6	11.1					
	Mycophenolat-Mophetil	1	1.9					
	Natalizumab	5	9.2					
	INF-β 1a s.c.	3	5.5					
	Rituximab	0	0					
	Teriflunomide	1	1.9					
Control Group matched	**Medication**	**n= 44**	**%**	**Medication**	**n= 44**	**%**	**Switched to**	**n= 7**
	None	21	47.7	None	23	52.3		
	INF-β 1a i.m.	0	0	INF-β 1a i.m.	0	0	None	
	INF-β 1b	13	29.6	INF-β 1b	12	27.3	None	2
	GA	1	2.3	GA	2	4.5		
	Methylprednisolone monthly i.v.	0	0	MP m IV	1	2.3		
	Mitoxantrone	6	13.6	Mitoxantrone	3	6.8	None/GA	1/1
	INF-β 1a s.c.	3	6.8	INF-β 1a s.c.	3	6.8	None/Mitox	2/1
	Rituximab	0	0	Rituximab	0	0		
Control Group all	**Medication**	**n= 59**	**%**	**Medication**	**n= 59**	**%**	**Switched to**	**n= 14**
	None	29	49.1	None	32	54.2	INF-β 1b /Mitox	1/1
	INF-β 1a i.m.	2	3.4	INF-β 1a i.m.	1	1.7	None	1
	INF-β 1b	17	28.8	INF-β 1b	16	27.1	None/Mitox	3/1
	GA	1	1.7	GA	1	1.7		
	Methylprednisolone monthly i.v.	0	0	MP m IV	1	1.7		
	Mitoxantrone	6	10.2	Mito-xantrone	4	6.8	None/GA	1/2
	INF-β 1a s.c.	4	6.8	INF-β 1a s.c.	4	6.8	None/Mitox	3/1
	Rituximab	0	0	Rituximab	0	0		

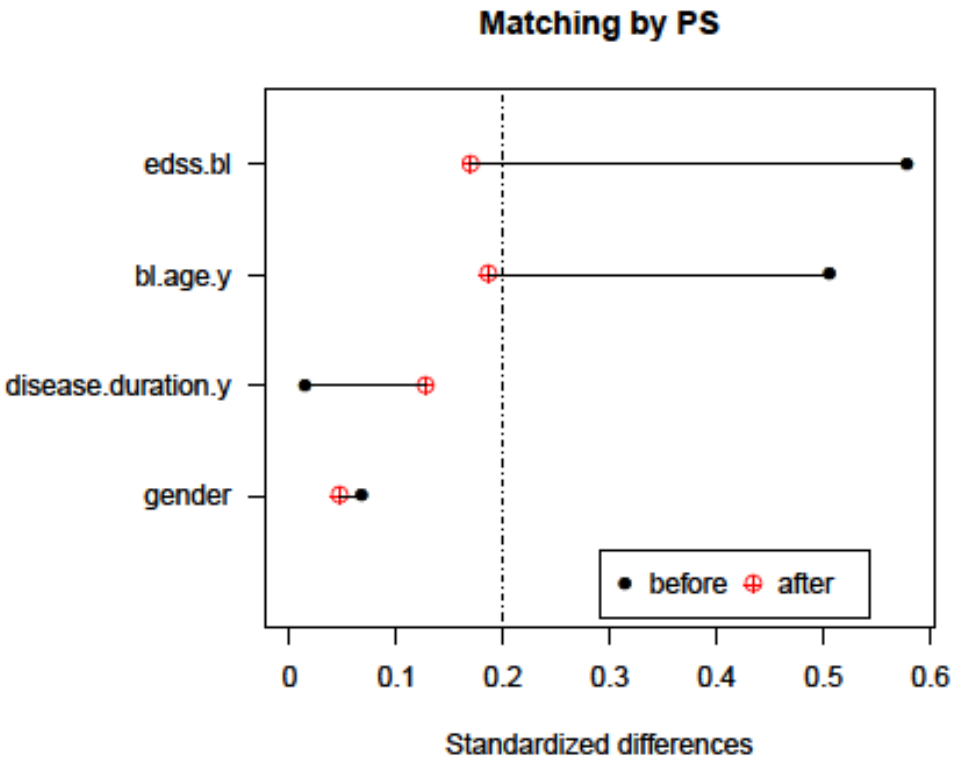
eTable 2: Treatment before BL and during follow-up for both cohorts (full and matched cohorts). “Switched to” indicates patients that switched during follow-up to another treatment. GA=Glatirameracetate

eFigure 1. Density Plot of Propensity Scores



eFigure 1: Density plot of the distribution of propensity scores in the rituximab and control group before matching (left) and after matching (right).

eFigure 2. Standardized Differences Between Groups



eFigure 2: Illustration of the standardized differences between groups in the total cohort and after propensity score (PS) matching (matched cohort).

RESEARCH

Association of delay of urgent or emergency surgery with mortality and use of health care resources: a propensity score–matched observational cohort study

Daniel I. McIsaac MD MPH, Karim Abdulla MD, Homer Yang MD, Sudhir Sundaresan MD, Paula Doering RN, Sandeep Green Vaswani MBA, Kednapa Thavorn MPharm PhD, Alan J. Forster MD MSc

Cite as: *CMAJ* 2017 July 10;189:E905-12. doi: 10.1503/cmaj.160576

See related article at www.cmaj.ca/lookup/doi/10.1503/cmaj.170172

ABSTRACT

BACKGROUND: Delay of surgery for hip fracture is associated with increased risk of morbidity and mortality, but the effects of surgical delays on mortality and resource use in the context of other emergency surgeries is poorly described. Our objective was to measure the independent association between delay of emergency surgery and in-hospital mortality, length of stay and costs.

METHODS: We identified all adult patients who underwent emergency noncardiac surgery between January 2012 and October 2014 at a single tertiary care centre. Delay of surgery was defined as the time from surgical booking to operating room entry exceeding institutionally defined acceptable wait times, based on a standardized 5-level priority system that accounted for surgery type and indication. Patients with delayed surgery were matched to those without delay using propensity scores derived from variables that accounted for details of admission and the hospital stay, patient characteristics, physiologic instability, and surgical urgency and risk.

RESULTS: Of 15 160 patients, 2820 (18.6%) experienced a delay. The mortality rates were 4.9% (138/2820) for those with delay and 3.2% (391/12 340) for those without delay (odds ratio [OR] 1.59, 95% confidence interval [CI] 1.30–1.93). Within the propensity-matched cohort, delay was significantly associated with mortality (OR 1.56, 95% CI 1.18–2.06), increased length of stay (incident rate ratio 1.07, 95% CI 1.01–1.11) and higher total costs (incident rate ratio 1.06, 95% CI 1.01–1.11).

INTERPRETATION: Delayed operating room access for emergency surgery was associated with increased risk of in-hospital mortality, longer length of stay and higher costs. System issues appeared to underlie most delays and must be addressed to improve the outcomes of emergency surgery.

Patients undergoing emergency surgery are at high risk of adverse outcomes.[1] Although patient characteristics[2,3] and surgical indication[4,5] are the most important risk factors, system factors, such as delayed access to the operating room, also affect outcomes. In hip fracture surgery, delay is associated with morbidity and mortality,[6,7] but for other surgeries, the effect of delay on outcomes is unclear.[8–13] Because it is very expensive to expand or reorganize operating room resources to improve access,[14–16] understanding the relation between delay and outcomes for all types of emergency surgery is needed.

The association between surgical delay and outcome may be obscured by confounding. The indication for surgery, comorbidities and physiologic disturbances may influence both the risk of delay and the risk of adverse outcomes. Furthermore, ascertainment of delay is a challenge. Many studies measure surgical wait time as the time from admission to surgery, but this is misleading, because inpatient work-up is often required to determine the risks and potential benefits of surgery.

The purpose of this study was to determine the independent association of surgical delay with inpatient mortality, postoperative length of stay and total costs of hospital care.

Methods

Design and setting

We performed a retrospective cohort study of emergency inpatient surgery at The Ottawa Hospital, a 900-bed academic health sciences centre serving a population of 1.2 million people. With 2 distinct campuses that perform emergency surgery, our institution is the regional cancer centre and the sole regional provider for trauma surgery, neurosurgery, thoracic surgery and vascular surgery. On weekdays, 5 or 6 of our 35 operating rooms are dedicated to emergency surgery; 3 of these rooms are available until 11 pm, and 2 are available between 11 pm to 7 am. All surgeons and anesthesiologists are paid on a fee-for-service basis.

Using established methods,[17] our hospital developed and adopted new wait time standards for emergency operating room access in January 2012 (described in detail in Appendix 1, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.160576/-/DC1). All emergency cases were classified into 1 of 5 urgency categories (A , < 45 min; B, < 2 h; C, < 4 h; D, < 8 h; E, < 24 h), according to the surgical procedure and patient indication.[18] The booking process was also clarified: patients could not be booked unless they were appropriately prepared to come to the operating room at any time after booking. The urgency classification was applied by the surgeon and was documented in the Surgical Information Management System. If there was a change in patient status (such as acute deterioration), the urgency classification could be changed accordingly.

Data sources

All data used in this study were derived from the data warehouse of The Ottawa Hospital, which stores clinical and administrative data. We used the Canadian Institute for Health Information Discharge Abstract Database; the National Ambulatory Care Reporting System; the Surgical Information Management System, which records all details of surgical procedures performed in our hospital operating rooms; and the electronic health record database at The Ottawa Hospital. Data sources are described in Appendix 2 (available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.160576/-/DC1).

Study population

We identified all adults undergoing emergency noncardiac surgery from January 2012 to October 2014. The study start date represents the implementation date of our 5-level prioritization system, and October 2014 was the last date at which all data sets were complete. We included all patients who were scheduled for emergency surgery and who entered the operating room within 3 multiples of the wait time for their assigned priority (representing the 95th percentile of wait times); we excluded patients who waited beyond this time frame because they may have been substantively different from those who waited less than 3 multiples of the accepted wait time (by virtue of their ability to survive this extended wait). All cases at The Ottawa Hospital are booked in the Surgical Information Management System as elective or emergent. Emergent cases are defined as those in which a patient is booked for surgery following evaluation and admission to hospital; elective cases are planned before admission. For our analysis, we included only cases in which the patient's clinical presentation necessitated operating room access within 24 hours.

Exposure

The exposure was surgical delay, defined as a wait time from surgical booking to the patient entering the operating room in excess of the accepted wait time for the patient's priority level. At the time of booking, each patient was assigned a priority level on the basis of our wait time standards (Appendix 3, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.160576/-/DC1). For our primary analysis, the exposure was treated as binary (delayed v. nondelayed). We classified each patient's urgency status according to the status in effect at the time of entry into the operating room.

When the acceptable wait time was exceeded, the operating room manager could submit 1 of 8 prepopulated reasons into the Surgical Information Management System to capture the reason for delay.

Outcomes

Our primary outcome was in-hospital mortality, the accuracy of which has been validated.[19] Length of stay was defined as the number of days from surgery to hospital discharge. Total hospital costs, calculated on a hospital-specific basis using standardized methods, included both direct and indirect costs standardized to 2014 Canadian dollars.[20] This method accounts for an individual patient's resource intensity weight and case-mix group, as well as fixed and indirect costs to the hospital based on the patient's location of care and length of stay.

Covariables

We specified covariables a priori on the basis of their likely role as confounders, through a structured literature review guided by an information specialist, the surgical risk calculator of the National Surgical Quality Improvement Program (American College of Surgeons)[3] and existing systematic reviews.[2] The full list of covariables and their representation in our models are available in Appendix 4 (available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.160576/-/DC1). We used several risk indices: Procedural Index for Mortality Risk, an internally validated index that predicts the risk of in-hospital death according to surgical procedure;[5] the Elixhauser Index,[21] an internally validated score that assigns points to comorbidities to predict in-hospital mortality;[22] and the Laboratory-Based Acute Physiology Score, an externally validated score that predicts physiology-associated mortality risk on the basis of laboratory values.[23]

Missing data

No outcome or exposure data were missing. Not all laboratory tests were ordered for all patients; we used published methods to impute missing values for the laboratory tests used in the Laboratory-Based Acute Physiology Score.[23]

Statistical analysis

We compared patient characteristics between groups using standardized differences, with differences of less than 10% being thought to represent negligible correlations.[24]

The primary analysis consisted of unadjusted and propensity score–matched analyses to measure the association between surgical delay and death. We used a nonparsimonious multivariable logistic regression model to estimate the propensity for delay for each patient. We included all variables postulated to act as confounders between delay and outcome (Appendix 4). We matched patients 1:1 without replacement based on a caliper (0.2 standard deviations of the propensity score logit) in a greedy matching algorithm (% gmatch).[25] Match quality was assessed in terms of the covariable balance between exposure groups and the achievement of a high proportion of matched exposed patients.[26] We measured the association between delay and mortality using a χ^2 test; the absolute risk difference and number needed to harm were also calculated. We tested an interaction between priority level and delay status to evaluate whether the effect of delay on mortality differed across priority levels.

For secondary analyses of length of stay and cost, we used the same matched cohort. Because the length of stay and cost distributions were skewed, we used nonparametric Wilcoxon signed-rank tests. We calculated relative associations using generalized linear models with log-link and gamma (cost) or negative binomial (length of stay) distributed errors.

We conducted several prespecified sensitivity analyses. We used the propensity score to assign inverse probability of treatment weights. This allowed us to calculate the average treatment effect, as opposed to the average treatment effect for the treated patients from the matched analysis.[27] We used prespecified inverse probability of treatment weight analyses based on subgroup-specific propensity scores to evaluate the robustness of our primary analysis with exclusion of hip fractures and exclusion of individuals delayed for medical reasons. We also performed a post hoc analysis restricted to patients with hip fracture.

Finally, we analyzed a continuous, but possibly nonlinear, relation between wait time and mortality, in contrast to the binary exposure used in our primary analysis. Wait times were standardized among priority levels by dividing the wait time in minutes by the accepted wait-time window. We used a generalized additive model, adjusting for procedural mortality risk (with the Procedural Index for Mortality Risk score) and patient comorbidity (with the Elixhauser Index) to analyze the continuous association between wait time and mortality. Splines with 3 degrees of freedom were used for each variable.

We used SAS version 9.4 (SAS Institute Inc.) for data management and analysis.

Ethics approval

This study was approved by the Ottawa Health Science Network Research Ethics Board.

Results

A total of 15 275 emergency surgery patients were identified (Figure 1). We excluded 115 patients because their wait times exceeded 3 multiples of the accepted wait-time window. Of the 15 160 included patients, 12 340 (81.4%) reached the operating room within an acceptable wait time, whereas 2820 (18.6%) did not (Table 1). The reasons for delay were documented in 1109 cases (39.3% of all delays; Table 2). Risk of mortality did not differ between patients with and without a documented reason for delay, and mean excess wait times are described in Appendix 5 (available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.160576/-/DC1). In 958 (86.4%) of the cases with a documented reason for delay, system issues such as availability of human or physical resources were identified as the reason for delay. The most common surgical procedures for each priority level are shown in Table 3.

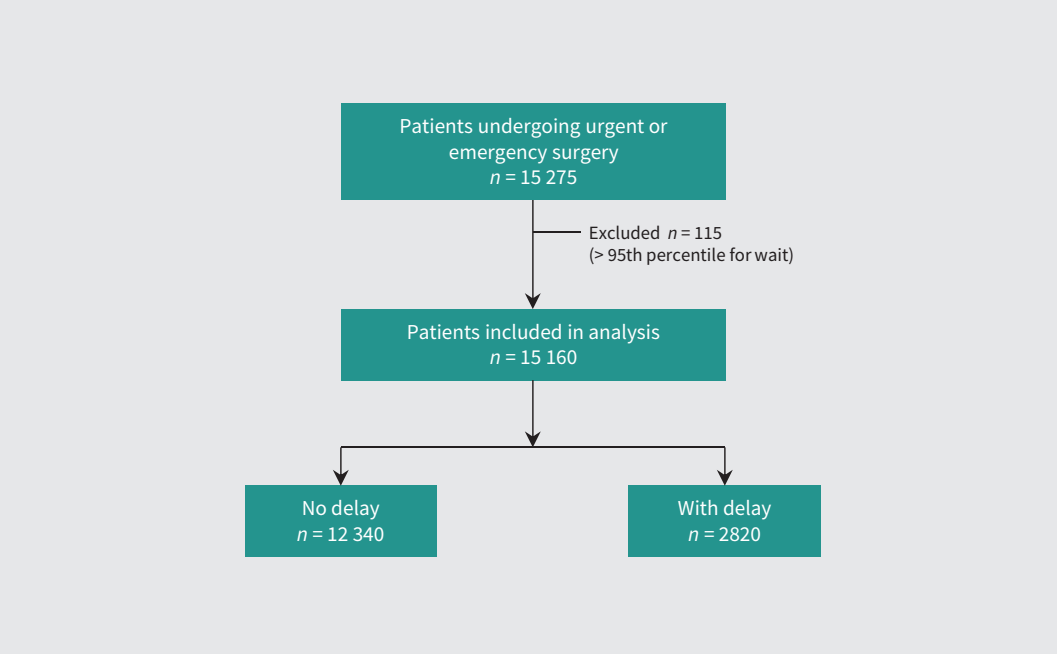


Figure 1: Study flow diagram.

Table 1: Demographic characteristics

Characteristic	% of entire cohort*			% of propensity score–matched cohort*		
	Nondelayed *n* = 12 340	Delayed *n* = 2820	Absolute standardized difference†	Nondelayed *n* = 2820	Delayed *n* = 2820	Absolute standardized difference†
Priority level						
A	4.6	7.4	11.8	6.4	7.4	3.9
B	7.5	5.4	8.6	5.4	5.4	0
C	10.2	7.1	16.7	7.3	7.1	0.8
D	29.9	12.2	44.5	12.0	12.2	0.6
E	47.9	67.9	41.4	68.9	67.9	2.2
Age, yr, mean ± SD	56.7 ± 20.7	60.8 ± 20.3	14.7	59.5 ± 20.8	60.8 ± 20.3	6.3
Sex, female	50.0	51.9	3.8	50.3	51.9	3.2
Elixhauser Index, mean (IQR)	2.0 (4.4)	2.0 (4.4)	0	1.9 (4.3)	2.1 (4.4)	4.6
Hospital admission status, urgent	95.2	95.6	1.9	95.7	95.6	0.5
Hospital entry through ED	84.8	84.7	0.3	84.6	84.7	0.3
ASA physical status score						
I and II	32.0	25.8	13.7	28.3	25.8	5.6
III and IV	65.7	71.8	13.2	69.7	71.8	4.6
V	2.2	2.3	0.7	1.9	2.3	2.8
Surgical service of record						
Dental	1.0	1.9	7.5	2.0	1.9	0.7
General surgery	32.3	21.1	25.5	21.1	21.1	0
Gynecology	5.0	2.6	12.6	2.4	2.6	1.3
Neurosurgery	7.2	8.2	3.8	8.6	8.2	1.4
Ophthalmology	1.5	1.2	2.6	1.3	1.2	0.9
Orthopedic surgery	27.3	42.2	31.7	41.9	42.2	0.6
Otolaryngology	2.0	1.9	0.7	1.8	1.9	0.7
Plastic surgery	1.6	0.8	7.4	0.8	0.8	0
Thoracic surgery	2.2	2.6	2.6	2.6	2.6	0
Urology	12.0	9.5	37.5	9.9	9.5	1.4
Vascular surgery	7.2	7.6	1.5	7.1	7.6	1.9
Other	0.8	0.1	3.7	0.1	0.1	0
Direct to operating theatre from ED	9.3	3.5	23.9	3.2	3.5	1.7
In ICU when surgery booked	3.2	3.5	1.7	3.2	3.5	1.7
PIMR score, mean ± SD	0.7 ± 1.5	0.9 ± 1.5	10.0	0.9 ± 1.6	0.9 ± 1.5	0
CTAS						
1	3.6	2.7	5.2	4.0	2.7	7.2
2	40.8	37.4	7.0	37.5	37.4	0.2
3	52.8	57.8	10.1	56.4	57.8	2.8
4–5	2.9	2.2	4.4	2.9	2.2	4.4
Transfusion within 24 h before surgery	3.7	5.5	8.6	4.3	5.5	5.6
Piperacillin–tazobactam or meropenem before arrival in operating room	6.6	8.2	6.1	7.4	8.2	3.0
Preoperative steroids	4.0	5.6	7.5	5.2	5.6	1.8
Preoperative insulin	11.5	14.9	10.1	13	14.9	5.5
LAPS, mean ± SD	22.6 ± 24.1	22.0 ± 37.2	1.9	22.4 ± 21.7	22.0 ± 37.2	3.2

Note: ASA = American Society of Anesthesiology physical status score (higher categories predict higher risk of death), CTAS = Canadian Emergency Department Triage and Acuity Scale, ED = emergency department, ICU = intensive care unit, IQR = interquartile range, LAPS = Laboratory Acute Physiology Score (range 0–256; higher scores predict higher risk of death), PIMR = Procedural Index for Mortality Risk, SD = standard deviation.
*Except where indicated otherwise.
†Absolute standardized differences > 10 are considered represent an important difference.

Crude mortality rate, length of stay and costs were higher in the group whose surgery was delayed (Table 4). We matched each of the 2820 delayed patients with a nondelayed patient using propensity scores, and achieved a balance of covariables (Table 1; Appendix 6, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.160576/-/DC1). Within the matched cohort, the risk of in-hospital mortality, length of stay and costs were all significantly higher in the delayed group (Table 4). The number needed to harm for mortality was 60; the odds ratio (OR) was 1.56 (95% confidence interval [CI] 1.18–2.06). An interaction term between priority level and delay status was not statistically significant ($p = 0.4$).

Sensitivity analyses

Inverse probability of treatment weight groups were balanced across covariables (Appendix 7, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.160576/-/DC1). In the full cohort, patients with delay were more likely to die (OR 1.35, 95% CI 1.08–1.69). When hip fractures were excluded, patients with delay had a 28% increase in relative odds of mortality (OR 1.28, 95% CI 1.00–1.65). When patient-specific delays were excluded, delay was associated with mortality (OR 1.51, 95% CI 1.14–2.00). Delay was also associated with mortality in the post hoc analysis of patients with hip fracture (OR 1.64, 95% CI 1.05–2.61). Appendix 8 (available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.160576/-/DC1) contains other post hoc analyses.

When we modelled the odds of mortality as a nonlinear continuous association with wait time, we found a significant, biphasic association ($p = 0.01$). The shape of the spline suggested that as the wait increased up to 1.1 multiples of the accepted wait-time window, the odds of mortality increased; beyond 1.1 multiples of the accepted wait-time window, the odds of mortality decreased. At 0.4 multiples of the accepted wait-time window, the odds of death crossed the null value (odds = 1), suggesting that beyond this point, excess mortality was attributable to increasing wait time (up to the 1.1 multiples of the accepted wait-time window inflection point; Figure 2).

Interpretation

Delay of operating room access for emergency surgery was independently associated with an increased risk of in-hospital mortality, longer length of stay and higher total hospital costs. Importantly, system factors, such as the availability of clinicians and physical resources, appeared to be the main reasons for delay of emergency surgery. We used validated indices to adjust for physiologic disturbance, preoperative therapies and timing of the decision to operate; this approach allowed us to address the limitations of prior research, which was unable to account for many acute preoperative patient variables and relied upon the time from hospital admission to operating room entry,[7] and to gain new insights into the implications of delay of surgery on use of health care resources.

Improving operating room access for patients needing emergency surgery will require careful consideration of the

Table 2: Reasons for delay in access ($n = 1109$)

Reason	No. (%) of patients
Availability of personnel	352 (31.7)
Anesthesiologist	42
Nurse	5
Surgeon	305
Availability of physical resources	147 (13.3)
Operating room	122
Postanesthesia care unit	11
Equipment	14
Multifactorial delay	459 (41.4)
Bumped by higher priority case	459
Patient-specific delay	151 (13.6)
Medically complex or decompensated patient	151

Table 3: Most common surgical procedures for each priority level

Priority level; reason and % of cases (by priority level)				
Priority A $n = 770$	**Priority B** $n = 1082$	**Priority C** $n = 1456$	**Priority D** $n = 4028$	**Priority E** $n = 7824$
Ruptured abdominal aortic aneurysm (9.6%)	Small-bowel repair or bypass for acute peritonitis (11.1%)	Hernia repair with bowel obstruction (9.0%)	Laparoscopic appendectomy (28.1%)	ORIF hip or neck of femur (27.5%)
Peripheral arterial bypass for acute distal ischemia (8.0%)	Large-bowel repair or bypass for acute peritonitis (10.2%)	Small-bowel repair or bypass for bowel obstruction (6.8%)	Laparoscopic cholecystectomy (4.8%)	Abdominal wall repair (8.2%)
Craniotomy for acute intracranial condition (7.6%)	Peripheral arterial bypass for vein graft occlusion (5.2%)	Large-bowel repair or bypass for bowel obstruction (5.4%)	Hernia repair (no peritonitis or bowel obstruction) (4.7%)	Ureteroscopy (5.6%)

Note: ORIF = open reduction internal fixation.

perioperative health care system, in particular how to optimize utilization of currently available resources and how to balance the increased resources that may be needed to improve access with the expected benefit of improved outcomes. We found that many delays were due to physician unavailability. Therefore, improving availability of personnel may improve access without increasing costs (at least from a hospital perspective). Furthermore, the increased resources required to have adequate nursing personnel and physical resources should be partially offset by decreased hospital costs independently attributed to surgical delay.

Considering the mechanisms underlying the delay–mortality relation, which in hip surgery may be attributable to hospital-acquired complications, might also help to properly target resources.[7] Unfortunately, our data do not contain validated measures of complications to support the generalizability of this mech-

Table 4: Study outcomes

Outcome	Nondelayed	Delayed	Absolute difference	Relative association* (95% CI)
Unadjusted	***n* = 12 340**	***n* = 2820**		
In-hospital death, no. (%)	391 (3.2)	138 (4.9)	1.6	1.59 (1.30–1.93)
Length of stay, d, mean ± SD	11 ± 19.1	13.6 ± 20.4	2.6	1.21 (1.16–1.25)
Total hospital costs,† $, mean ± SD	19 144 ± 33 900	22 479 ± 40 224	3335	1.17 (1.13–1.21)
Propensity score match adjusted	***n* = 2820**	***n* = 2820**		
In-hospital death, no. (%)	90 (3.2)	138 (4.9)	1.7	1.56 (1.18–2.06)
Length of stay, d, mean ± SD	12.5 ± 20.7	13.6 ± 20.4	1.1	1.07 (1.01–1.11)
Total hospital costs,† $, mean ± SD	20 989 ± 35 085	22 479 ± 40 224	1490	1.06 (1.01–1.11)

Note: CI = confidence interval, SD = standard deviation,
*For death , relative association is odds ratio; for length of stay and costs, relative association is the incidence rate ratio.
†All costs are in 2014 Canadian dollars.

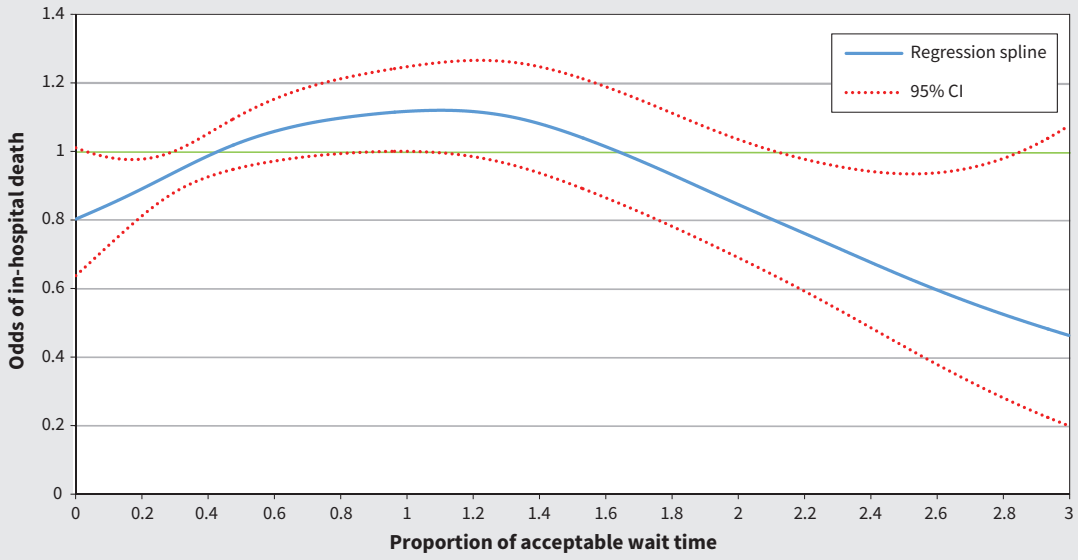


Figure 2: Spline of risk of in-hospital death as a function of wait time. The adjusted regression spline presents the proportion of acceptable wait time experienced by emergency surgical patients as a function of the odds of in-hospital death. Where the spline and 95% confidence limits are greater than the green line at 1, there is a significantly increased risk of death. CI = confidence interval.

anism. However, we did find a nonlinear association between wait time and mortality, which suggests opportunities to match patient need for rapid operating room access with scarce resources. The odds of dying increased for waits of up to 1.1 multiples of the accepted wait-time window, and decreased thereafter. This relation suggests that there may be 2 subgroups of patients. One group truly appears to need urgent surgical intervention; in these patients, longer waits may increase the risk of death. Decreasing mortality risk beyond 1.1 wait-time multiples may indicate a group experiencing survivorship bias, whose outcome is not as sensitive to longer wait times for surgery. Therefore, a more granular approach, considering both patient and procedural characteristics, is needed to appropriately map wait-time standards to tolerance for waiting. Because the odds of death crossed the null value (odds = 1) at 0.4 wait-time multiples, the wait times established by our hospital may be too long. Prospective research that provides a clear understanding of the interaction between patient and procedural risk is needed to inform stakeholders in the application of scarce health care resources aimed at improving emergency surgery outcomes through more timely operating room access. Our findings that many emergency surgical cases are delayed, that delays are typically due to system issues and that delays are associated with adverse outcomes are consistent with the existing literature. In the current study, about 20% of patients needing emergency surgery did not gain access to the operating room in a time frame compliant with our institutional guidelines, and 86.4% of delays were attributable to system factors, similar to the findings of previous studies.[28–30] In the United Kingdom, only 1% of emergency surgeries may be delayed because patient resuscitation is needed, and the majority of cases with delay experience organizational problems.[29] Among emergency laparotomies, 77% of delays were related to being bumped by higher-priority cases or unavailability of operating rooms.[28]

Although our results support the known association between delay and mortality in patients with hip fracture,[7] they also support the generalizability of the delay–outcome association to nonorthopedic surgery, where previous findings have both supported[8,9] and refuted[10,11] this association. There are about 100 000 emergency general surgery operations,[31] and 30 000 hip fracture operations annually in Canada.[32] If 20% of these cases experience delays, we estimate that more than 410 deaths may be attributable to surgical delay each year. A randomized trial of timing of hip fracture surgery is ongoing,[33] but randomized data are also needed to evaluate the effect of timing of emergency surgery for other indications.

Limitations

A causal and generalizable relation between delay of surgery and death cannot be determined from a single-centre observational study; a randomized trial would be required. Despite our use of robust methods, confounding and unmeasured covariables are threats to the validity of our results. Our primary classification of delay was dichotomization of a continuous variable; our sensitivity analysis based on a generalized additive model was a secondary analysis, and must be interpreted as such. Our classification of delay was based on a prioritization system that has not been tested with regard to its effect on outcomes. The reported reasons for noncompliance with timing of emergency surgery were missing for about 60% of patients. Finally, we cannot account for deaths that may have occurred outside our institution.

Conclusion

Delay in operating room access for emergency surgery is associated with increases in mortality risk, length of stay and costs. These findings are consistent with increasing evidence showing that delays in emergency surgery cause harm. Improving timely access to emergency surgery may require reallocation of scarce resources; however, such reallocation may be offset by savings derived from avoiding delays.

References

1. Haider AH, Obirieze A, Velopulos CG, et al. Incremental cost of emergency versus elective surgery. *Ann Surg* 2015;262:260-6.
2. Davis P, Hayden J, Springer J, et al. Prognostic factors for morbidity and mortality in elderly patients undergoing acute gastrointestinal surgery: a systematic review. *Can J Surg* 2014;57:E44-52.
3. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg* 2013;217:833-42.e1-3.
4. Fleisher LA, Beckman JA, Brown KA, et al.; American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines on Perioperative Cardiovascular Evaluation for Noncardiac Surgery); American Society of Echocardiography; American Society of Nuclear Cardiology; Heart Rhythm Society; Society of Cardiovascular Anesthesiologists; Society for Cardiovascular Angiography and Interventions; Society for Vascular Medicine and Biology; Society for Vascular Surgery. ACC/AHA 2007 guidelines on perioperative cardiovascular evaluation and care for noncardiac surgery: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 2002 Guidelines on Perioperative Cardiovascular Evaluation for Noncardiac Surgery). *Anesth Analg* 2008;106:685-712.
5. van Walraven C, Wong J, Bennett C, et al. The Procedural Index for Mortality Risk (PIMR): an index calculated using administrative data to quantify the independent influence of procedures on risk of hospital death. *BMC Health Serv Res* 2011;11:258.
6. Moja L, Piatti A, Pecoraro V, et al. Timing matters in hip fracture surgery: patients operated within 48 hours have better outcomes. A meta-analysis and meta-regression of over 190,000 patients. *PLoS One* 2012;7:e46175.
7. Simunovic N, Devereaux PJ, Sprague S, et al. Effect of early surgery after hip fracture on mortality and complications: systematic review and meta-analysis. *CMAJ* 2010;182:1609-16.
8. Buck DL, Vester-Andersen M, Møller MH. Surgical delay is a critical determinant of survival in perforated peptic ulcer. *Br J Surg* 2013;100:1045-9.
9. Azuhata T, Kinoshita K, Kawano D, et al. Time from admission to initiation of surgery for source control is a critical determinant of survival in patients with gastrointestinal perforation with associated septic shock. *Crit Care* 2014;18:R87.
10. Vester-Andersen M, Lundstrøm LH, Buck DL, et al. Association between surgical delay and survival in high-risk emergency abdominal surgery. A population-based Danish cohort study. *Scand J Gastroenterol* 2016;51:121-8.
11. Sicard N, Tousignant P, Pineault R. Non-patient factors related to rates of ruptured appendicitis. *Br J Surg* 2007;94:214-21.
12. Ong M, Guang TY, Yang TK. Impact of surgical delay on outcomes in elderly patients undergoing emergency surgery: a single center experience. *World J Gastrointest Surg* 2015;7:208-13.
13. Ingraham AM, Cohen ME, Bilimoria KY, et al. Effect of delay to operation on outcomes in adults with acute appendicitis. *Arch Surg* 2010;145:886-92.
14. Anantha RV, Paskar D, Vogt K, et al. Allocating operating room resources to an acute care surgery service does not affect wait-times for elective cancer surgeries: a retrospective cohort study. *World J Emerg Surg* 2014;9:21.
15. Heng M, Wright JG. Dedicated operating room for emergency surgery improves access and efficiency. *Can J Surg* 2013;56:167-74.
16. Hameed SM, Brenneman FD, Ball CG, et al.; Canadian Association of General Surgery Committee on Acute Surgery and Critical Care. General surgery 2.0: the emergence of acute care surgery in Canada. *Can J Surg* 2010;53:79-83.

17. Smith CD, Spackman T, Brommer K, et al. Re-engineering the operating room using variability methodology to improve health care value. *J Am Coll Surg* 2013;216:559-68.
18. Ryckman FC, Adler E, Anneken AM, et al. Cincinnati Children's Hospital Medical Center: redesigning perioperative flow using operations management tools to improve access and safety.. In: Litvak E, editor. *Managing patient flow in hospitals: strategies and solutions*. 2nd ed. Oak Brook (IL): Joint Commission Resources; 2009:97-111.
19. Juurlink D, Preyra C, Croxford R, et al. Canadian Institute for Health Information Discharge Abstract Database: a validation study [ICES investigative report]. Toronto: Institute for Clinical Evaluative Sciences; 2006.
20. Wodchis WP, Bushmeneva K, Nikitovic M, et al. Guidelines on person-level costing using administrative databases in Ontario. Working Paper Series Vol. 1. Toronto: University of Toronto, Institute of Health Policy, Management and Evaluation, Health System Performance Research Network; 2013. Available: www.hsprn.ca/uploads/files/Guidelines_on_PersonLevel_Costing_May_2013.pdf (accessed 2016 Sept. 19).
21. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130-9.
22. van Walraven C, Austin PC, Jennings A, et al. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care* 2009;47:626-33.
23. Escobar GJ, Greene JD, Scheirer P, et al. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. *Med Care* 2008;46:232-9.
24. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Simul Comput* 2009;38:1228-34.
25. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008;27:2037-49.
26. Ali MS, Groenwold RHH, Belitser SV, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *J Clin Epidemiol* 2015;68:112-21.
27. Austin PC, Laupacis A. A tutorial on methods to estimating clinically and policy-meaningful measures of treatment effects in prospective observational studies: a review. *Int J Biostat* 2011;7:6.
28. Schneider C, Tyler LE, Scull EF, et al. A case-control study investigating factors of preoperative delay in emergency laparotomy. *Int J Surg* 2015;22:131-5.
29. Pearse RM, Dana EC, Lanigan CJ, et al. Organisational failures in urgent and emergency surgery. A potential peri-operative risk factor. *Anaesthesia* 2001;56:684-9.
30. Katsanis N, Rosenberg SM. An age-old problem. *PLoS Genet* 2007;3:e37.
31. Gale SC, Shafi S, Dombrovskiy VY, et al. The public health burden of emergency general surgery in the United States: a 10-year analysis of the Nationwide Inpatient Sample — 2001 to 2010. *J Trauma Acute Care Surg* 2014;77:202-8.
32. *Health indicators 2007*. Ottawa: Canadian Institute for Health Information; 2007.
33. Hip Fracture Accelerated Surgical Treatment and Care Track (HIP ATTACK) Investigators. Accelerated care versus standard care among patients with hip fracture: the HIP ATTACK pilot trial. *CMAJ* 2014;186:E52-60.

Competing interests: None declared.

This article has been peer reviewed.

Affiliations: Departments of Anesthesiology (McIsaac, Abdulla, Yang), Surgery (Sundaresan) and Medicine (Forster), University of Ottawa; The Ottawa Hospital (McIsaac, Yang, Sundaresan, Doering, Forster), Civic Campus; Ottawa Hospital Research Institute (McIsaac, Thavorn, Forster), Ottawa, Ont.; Institute for Healthcare Optimization (Green Vaswani), Newton, MA

Contributors: All of the authors contributed to study conception, and Daniel McIsaac, Karim Abdulla and Alan Forster contributed to study design. Daniel McIsaac, Karim Abdulla, Kednapa Thavorn and Alan Forster contributed to the analysis, and all of the authors contributed to interpretation of the data. Daniel McIsaac wrote the manuscript, with contributions from Karim Abdulla and Alan Forster. All of the authors reviewed the manuscript for important intellectual content, approved the final version for publication and agreed to act as guarantors of the work.

Funding: Daniel McIsaac receives salary support from The Ottawa Hospital Department of Anesthesiology. No other funding was received.

Acknowledgements: This project was undertaken in collaboration with the Institute for Healthcare Optimization (Boston, MA; ihoptimize.org). The authors thank Jocelyn Tufts for assistance in creating the analytical data set.

Accepted: Feb. 6, 2017

Correspondence to: Daniel McIsaac, dmcisaac@toh.on.ca

Appendix 4 (as supplied by the authors): Propensity Score Model for full cohort

Abbreviations-ASA-American Society of Anesthesiologists; CI: confidence interval; CTAS: Canadian triage and acuity scale; ED: emergency department; ENT: ear nose and throat (otolaryngology); ICU: intensive care unit; LAPS: laboratory acute physiology score; LOS: length of stay; PIMR: procedural index for mortality risk; NA: not applicable; OR: odds ratio for surgical delay

Covariate	Representation	OR (95%CI)
Preoperative LOS	< 24 hrs	Referent (NA)
	24-72 hrs	0.98 (0.83-1.15)
	>72 hrs	0.76 (0.68-0.86)
Priority	A	Referent (NA)
	B	0.34 (0.26-0.44)
	C	0.28 (0.22-0.36)
	D	0.15 (0.12-0.19)
	E	0.38 (0.30-0.48)
Age	<65 years	Referent (NA)
	65-74 year	1.09 (0.96-1.24)
	>75 years	1.05 (0.93-1.18)
Male	vs Female	0.9 (0.82-0.99)
Congestive heart failure		1.17 (0.87-1.57)
Cardiac arrythmias		0.84 (0.69-1.03)
Valvular disease		1.59 (1.01-2.51)
Pulmonary circulation disorders		1.21 (0.80-1.83)
Peripheral vascular disorders		0.92 (0.71-1.21)
Hypertension, uncomplicated		1.09 (0.95-1.25)
Paralysis		1.04 (0.69-1.56)
Other neurological disorders		0.84 (0.60-1.17)
Chronic obstructive pulmonary disease		0.93 (0.71-1.22)
Diabetes, uncomplicated		0.91 (0.73-1.12)
Diabetes, complicated		0.9 (0.74-1.09)
Hypothyroidism		1.04 (0.68-1.61)
Renal failure		1.42 (1.06-1.91)
Liver disease		1.36 (0.86-2.15)
Peptic ulcer disease		1.48 (0.50-4.33)
AIDS/HIV		1.29 (0.45-3.66)
Lymphoma		0.7 (0.42-1.16)
Solid tumor without metastases		0.88 (0.72-1.07)
Metastatic cancer		0.98 (0.77-1.26)
Connective tissue diseases		0.86 (0.52-1.41)
Coagulopathy		0.53 (0.33-0.86)
Obesity		0.78 (0.48-1.26)
Weight loss		0.59 (0.48-1.44)
Fluid and Electrolyte Disorders		0.89 (0.64-1.22)
Blood loss anemia		3.85 (1.54-9.61)
Deficiency anemia		1.03 (0.51-2.10)
Alcohol abuse		1.28 (0.87-1.89)

Appendix to: McIsaac DI, Abdulla K, Yang H, et al. Association of delay of urgent or emergency surgery with mortality and use of health care resources: a propensity score–matched observational cohort study. *CMAJ* 2017. DOI:10.1503/cmaj.160576.

Drug abuse			1.83 (1.03-3.27)
Psychoses			0.67 (0.34-1.33)
Depression			0.86 (0.55-1.34)
Elixhauser Index quartile		0	Referent (NA)
		1	0.84 (0.50-1.41)
		2	1.03 (0.36-2.91)
		3	0.94 (0.54-1.64)
Hospital Admission Status - Urgent	vs Elective		1.16 (0.90-1.50)
Hospital Entry	via Clinic		Referent (NA)
	Direct to service		0.89 (0.59-1.36)
	via Emergency		1.36 (0.65-2.88)
	From day procedure		0.55 (0.19-1.54)
ASA Score	I and II		Referent (NA)
	III and IV		1.05 (0.94-1.18)
	V		0.74 (0.52-1.06)
Other			Referent (NA)
Dental			2.7 (1.22-6.03)
General Surgery			0.93 (0.45-1.93)
Gynecology			0.69 (0.32-1.48)
Neurosurgery			1.33 (0.63-2.79)
Ophthalmology			1.76 (0.77-4.03)
Orthopedic Surgery			1.68 (0.81-3.48)
ENT			1.02 (0.47-2.22)
Plastic Surgery			0.65 (0.28-1.51)
Thoracic Surgery			1.33 (0.62-2.87)
Urology			0.91 (0.43-1.89)
Vascular Surgery			1.05 (0.50-2.23)
Direct to OR from the Emergency Dept.			0.34 (0.27-0.43)
In ICU when Surgery booked			0.79 (0.61-1.03)
PIMR Score quartile		0	Referent (NA)
		1	1.1 (0.63-1.91)
		2	1.09 (0.97-1.22)
		3	1.03 (0.81-1.31)
CTAS		1	Referent (NA)
		2	0.69 (0.36-1.31)
		3	0.74 (0.39-1.41)
		4 or 5	0.65 (0.20-2.11)
Transfusion within 24 hours of surgery			0.61 (0.41-0.92)
Pip/Tazo or Meropenem on arrival to OR			1.29 (1.08-1.53)
Preop Steroids			1.27 (1.04-1.56)
Preop Insulin (%)			1.14 (0.95-1.37)
LAPS		0	Referent (NA)
		1	1.08 (0.94-1.23)
		2	1.12 (0.97-1.29)
		3	0.98 (0.84-1.14)

Appendix to: McIsaac DI, Abdulla K, Yang H, et al. Association of delay of urgent or emergency surgery with mortality and use of health care resources: a propensity score–matched observational cohort study. *CMAJ* 2017. DOI:10.1503/cmaj.160576.

Macrothink Institute™

Journal of Agricultural Studies
ISSN 2166-0379
2015, Vol. 3, No. 2

Analysis of Effects of Agriculture Intervention Using Propensity Score Matching

Peter Josephat (Corresponding author)

Dept. of Statistics, University of Dodoma

P .O. Box 338, Dodoma, Tanzania

Tel: 255-787-288-998 E-mail: mtakwimu@yahoo.com

Rose Likangaga

Local Government Training Institute (LGTI)

P. O. Box 1125, Dodoma, Tanzania

Tel: 255-784-505-491 E-mail: likangagar@yahoo.com

Received: March 29, 2015 Accepted: April 24, 2015

doi:10.5296/jas.v3i2.7339 URL: http://dx.doi.org/10.5296/jas.v3i2.7339

Abstract

Nowadays, the agriculture extension programmes are practiced in many parts of the world. There is a mixture of results about the effects of agriculture intervention programmes. The literature shows that the interventions are ineffective and have limited diffusion. On the other side, it shows that interventions are effective. Following different arguments about the effects of agriculture extension, this paper adopted Propensity Score Matching (PSM) to analyze the effects of District Agricultural Sector Investment Project (DASIP) using agriculture data.

The study was conducted in rural Tanzania areas. It covered five regions namely Kagera, Mwanza, Mara, Simiyu and Kigoma. The study focused on agro-ecological zone where corn is cultivated. Two methods which are questionnaire administration and direct oral interview were used to collect primary data. The collection of data using the questionnaire was done from both participants (359) and non-participants (519). Before running the independent t test, the estimation of propensity score was done using Logistic regression. Thirteen confounding variables were used to estimate propensity scores.

The effects of the intervention were analysed by considering four items namely the earnings from corn production, value of livestock owned, value of household assets owned, and value of farm assets owned. The results show that none of the four factors had significant result as the p values are greater than 0.05. This implies that the earning between farmers participating in DASIP are not significant different from those who do not participate in the programme. The study recommends that the group activities should last longer rather than changing them from time to time.

Keywords: Agriculture Extension programme, DASIP, Farmer Field School, Intervention, Propensity Score Matching (PSM)

1. Introduction

The agriculture plays a major role in economic development (Yeshwanth, 2008). Nowadays, agriculture extension programmes are practiced in many parts of the world. Such programmes are implemented because farmers lack direct linkage with advanced agricultural technology. It is through extensions where farmers are given knowledge, skills and motivation for farming. These are done through Farmer Field Schools (FFS) also called Participatory Group Farmers (PGFs) model.

The FFS started in Tanzania in the 1997 (Braun et al., 2006). The approach has been engineered by both government and non-governmental organizations. The government of Tanzania adopted the FFS approach in one of its project called District Agricultural Sector Investment Project (DASIP) in which this paper is focused. The DASIP is a six year project aimed at increasing the productivity and incomes of rural households in the project area within the overall framework of the Agricultural Sector Development Strategy (ASDS). The DASIP started in the 2006.

One of the main challenges that the extension and research is currently confronted with is the transfer of agricultural technology from the research stations to the farm lands (Dinpanah et al., 2010). There is a mixture of results about the effects of agriculture intervention programmes. The literature shows that the interventions are ineffective and have limited diffusion (see Quizon et al., 2001; Feder et al., 2003; and Rola et al., 2002). On the other side, the literature shows that FFS are effective (see Godtland et al., 2004; Van den Berg., 2004; Feder et al., 2003; Tripp et al, 2005; Erickson, 2003; and Ooi et al, 2005).

There is less common rigorous impact evaluations of agricultural extension interventions despite the vast literature dealing with issues related to agricultural extension (Waddington et al., 2010). Heinrich et al. (2010) argue that this is a result of several problems accompanied by the evaluation of the programmes. The problems include: establishing the counterfactual; need for an adequate comparison group; selection bias; and role of randomization (Duflo and Kremer, 2003). These problems can be solved by the use of statistical methods depending on the nature of the intervention programmes. Unfortunately, the data used in the past impact analyses did not define well the counterfactual factors. The comparison is done by just looking at two observation points that is, before and after.

The intervention programmes can either be random or non-random. The randomized design

occurs when the inclusion of the units or subjects in the intervention is random while non-random implies that the inclusion of the units is not by chance but depends on other factors. The randomized design programmes are very limited in literature because many intervention programmes have specific objectives and target something which makes non-random inclusion of the units. Because of this, non-experimental methods are adopted.

The Propensity Score Matching (PSM) is among the vibrant non-experimental methods. It is the most used method because it overcomes the fundamental evaluation problem and addresses the possible occurrence of selection bias. One of the advantages of this method is that it is used even if there are no baseline data. This was argued by Caliendo and Kopeinig (2005) in their working paper about guidance for the implementation of PSM.

Following different arguments about the effects of agriculture extension, this paper adopted PSM to analyse effects of DASIP using agriculture data.

2. Material and Methods

2.1 Study Area

The study was conducted in rural Tanzania particularly in the areas where DASIP operates. This covers five regions namely Kagera, Mwanza, Mara, Simiyu and Kigoma. Within these regions, the study focused on agro-ecological zone where corn is cultivated. The rationale of choosing DASIP area was that there was limited statistical survey on the impact evaluation of the programme conducted.

2.2 Population and Sample Size

The target population of this study was corn farmers found in DASIP intervention areas. Both DASIP participants (242,000 farmers) and non-participants were included in the study. A sampling unit was individual farmer.

Basing on the sample selection formula by Yamane (1967), out of 242,000 participant farmers, only 399 participant farmers with the precision level (margin of error) of 5% were targeted in the sample but the study was able to include only 359 (89.97%) participant farmers. In order to control the confounding factors, the study involved the sample of 275 (96.2%) out of 286 non-participant farmers who were located in the villages with DASIP intervention and sample of 244 (85.3%) out of targeted 286 non-participant farmers located in the villages far from villages where DASIP operates (See Table 1).

Table 1. Sample Distribution

Corn farmer category	Targeted sample size	Actual sample size
FFS participant farmers from villages with FFS programme	399	359
Non-participant farmers from villages with FFS programme	286	275
Non-participant farmers from villages without FFS programme	286	244
Total	971	878

There was a need to have a large sample size of non-participant farmers in evaluating the

impact of a programme especially PSM (Bryson et al., 2002). The PSM requires large data in both the number of variables and sample size. It is termed as a data hungry method. When data are scarce, the appropriateness of this technique should be carefully analysed (Heinrich et al., 2010). For instance, the sample size used by Lalonde (1986) consisted of 297 treatment group observation and 425 control group (Smith and Todd, 2005). Actually, the large sample makes fewer margins of error and increases confident in results.

The systematic sampling with non-equal selection probabilities was involved in the study to select the respondents. The technique was applied because the population under study was not homogenous something which restricted the use of systematic sampling with equal selection probabilities.

2.3 Data Collection and Analysis

The two methods namely the questionnaire administration and direct oral interview were used to collect primary data. The collection of data using the questionnaire was done from both participant (359) and non-participant farmers (519). The use of the questionnaire was preferred in this study because it is very cost-effective and it gives a greater response rate rather than adopting inadequate mailing method and/or time consuming direct oral interview in such a large sample and large geographical area. This study adopted a pre-test structured questions form of interview for safe basis generalization. The structured interview method was applied to DASIP officers.

The analysis of the effects of extension was done using an independent t test. The effects of the intervention is analysed by considering four items namely the earnings from corn production, value of livestock owned, value of household assets owned, and value of farm assets owned. The earning from corn production refers the money obtained after selling corns. The value of the livestock such as chicken, guinea fowl, duck, etc. owned by the farmer comprised of the total value of all livestock. . The household assets include the total value of things such as house, mobile phone, television, radio, bicycle, etc. The farm assets include assets such as hand hoe, farm (ha), cart/barrow, plough, etc.

Before running the t test, the estimation of propensity score was done using Logistic regression. Thirteen confounding variables involved in estimation included: sex, age, type of farmer, marital status, level of education, household size, land owned, distance from home to corn farm (km), distance from the village to district headquarters (km), distance to tarmac road (km), weather, soil type and membership of other participatory farmer groups.

3. Results and Discussion

3.1 Propensity Score Matching

Before the PSM was performed, the data were scrutinized to ensure that they are clean. The variables with outlier were cleaned. The estimated propensity score were matched using simple 1:1 nearest neighbour matching. The calliper of 0.15 of the standard deviation of the logit of the propensity score was used to exclude bad matches as recommended by Thoemmes (2012).

In order to assess the balance achieved after matching, both univariate and multivariate balance statistics are used. For multivariate, chi square and $\mathcal{L}_1$ were used while for univariate standardized mean difference $|d|$ and plots were used.

The output shows that the overall chi square balance test was not significant as $\chi^2(13) = 14.878, p = 0.315$. Multivariate imbalance measure $\mathcal{L}_1$ for unmatched solution (before matching) was 0.998 while after matching was 0.996. Both Chi square test and multivariate imbalances show that there is no imbalance after matching. The $\mathcal{L}_1$ indicates that there is no imbalance because the value for matched sample is small (0.996) than unmatched sample (0.998).

For the case of univariate balance test, the standardized mean difference shows that all covariates were balanced as $|d| \leq 0.25$. The distribution of propensity scores is shown in Figure 1. The graph is visualized to examine the similarity of the propensity score distributions after matching and to assess the area of common support (Thoemmes, 2012). From Figure 1, it can be seen that there is overlapping distribution of the propensity scores in treatment and control groups as tails of histograms are overlapping.

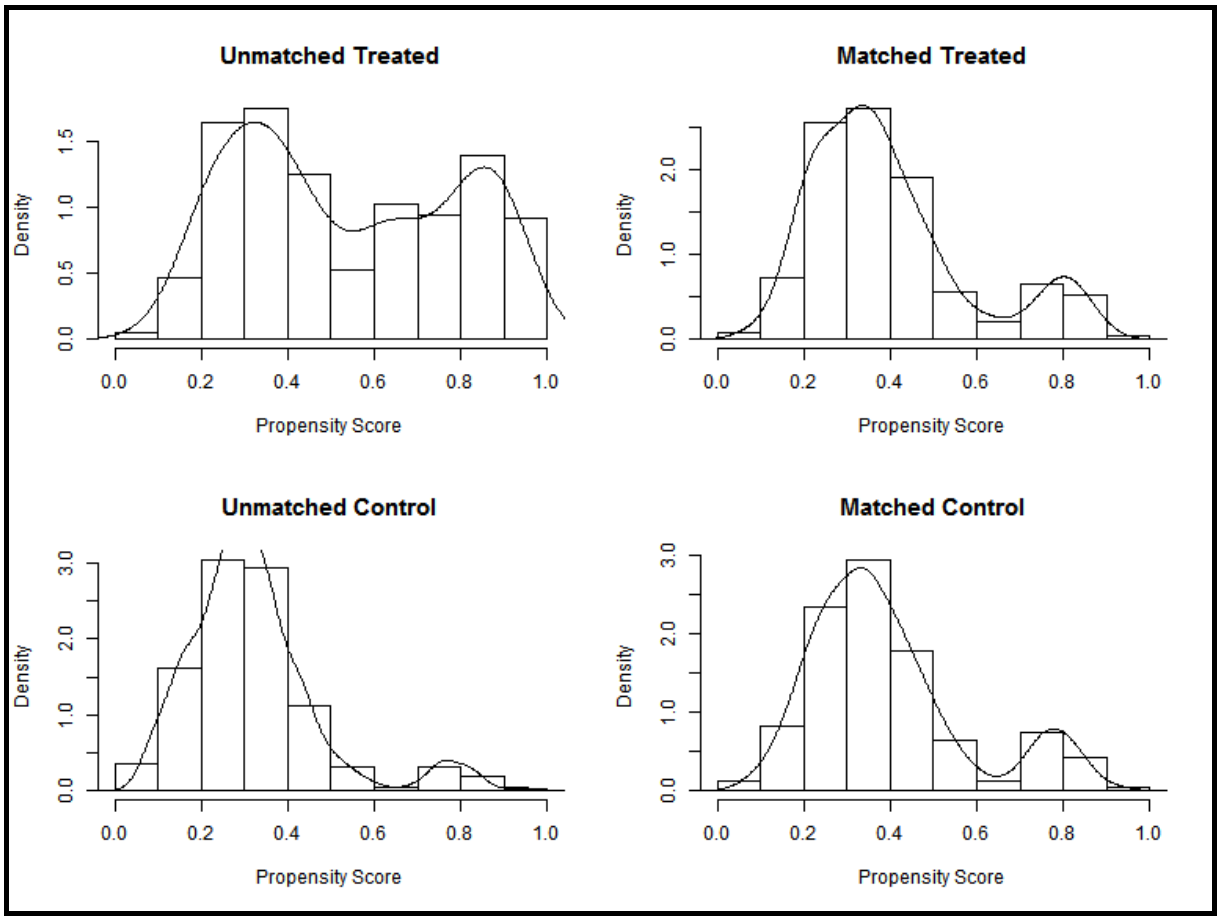


Figure 1. Distribution of Propensity Scores of Treatment and Control Groups

In Figure 2, the plot showing individual propensity scores is presented. Unlike plot in Figure 1, Figure 2 provides information of plots of individual units. It can be easily seen that the two groups (treatment and control) are comparable.

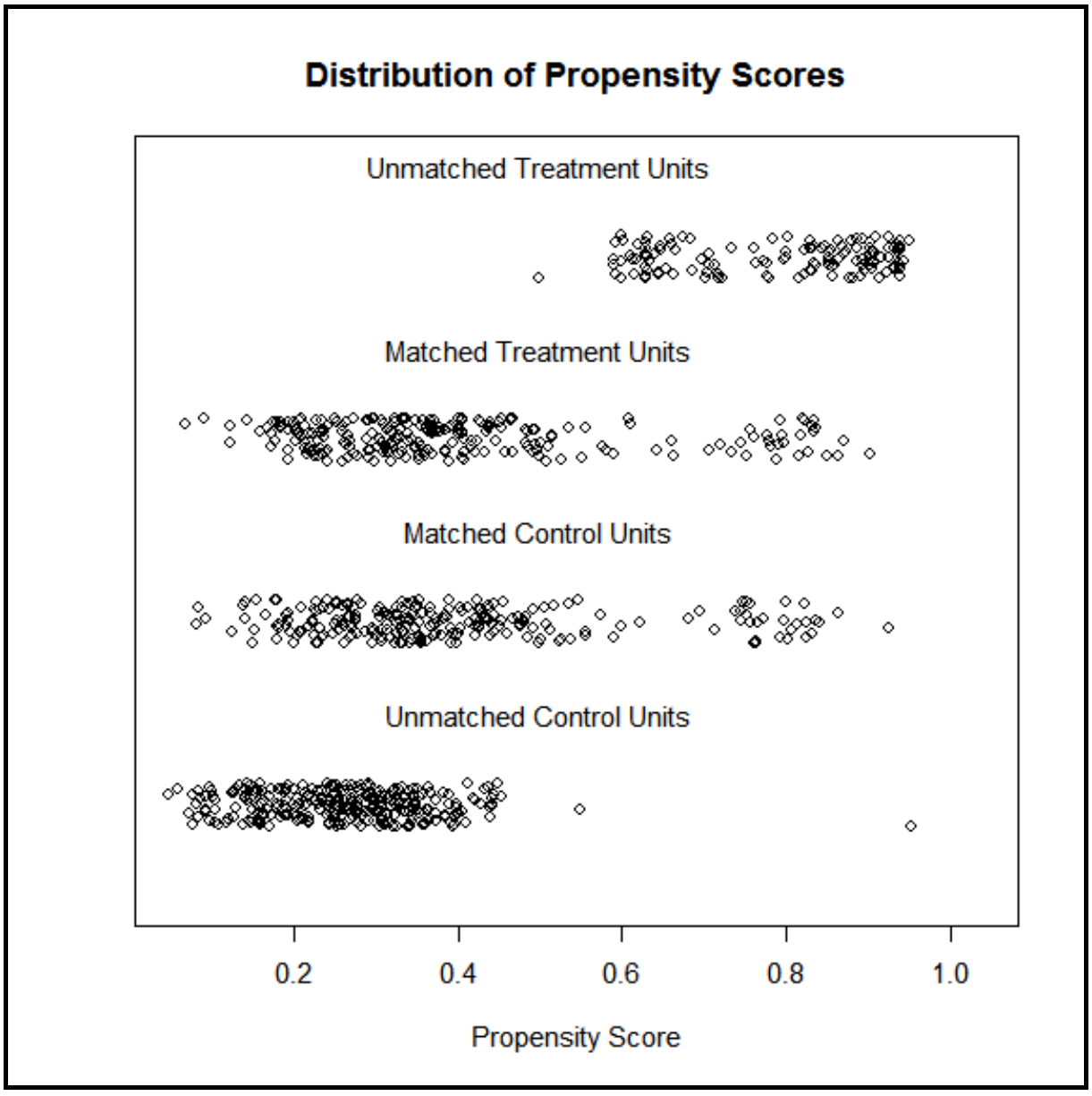


Figure 2. Dot plot of Individual Propensity Scores

The standardized difference is presented in Figure 3. From the figure, it can be seen that the standardized differences after matching are centred on zero and that no systematic difference still exist after matching.

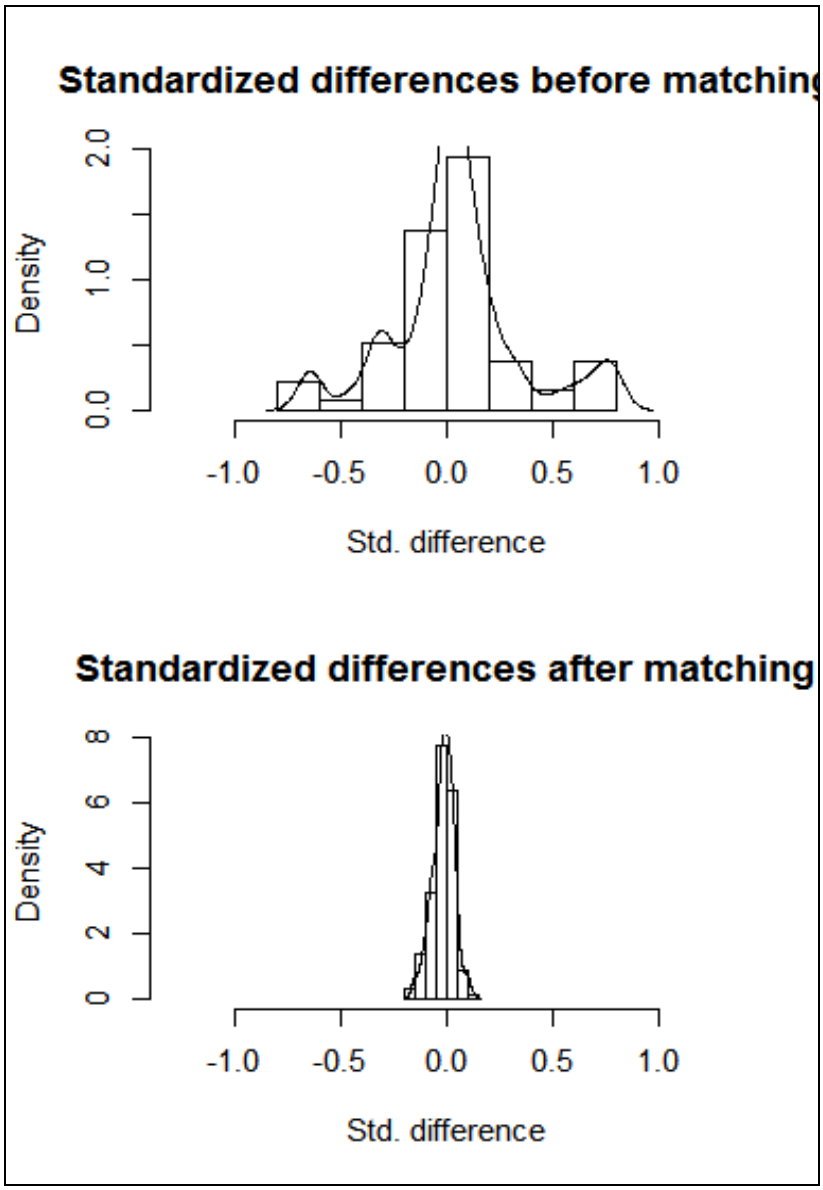


Figure 3. Histogram of Standardized Differences

The magnitude of standardized differences before and after matching for each covariate is presented in Figure 4. It can be seen from the figure that there is an improvement of scores after matching compared to before matching.

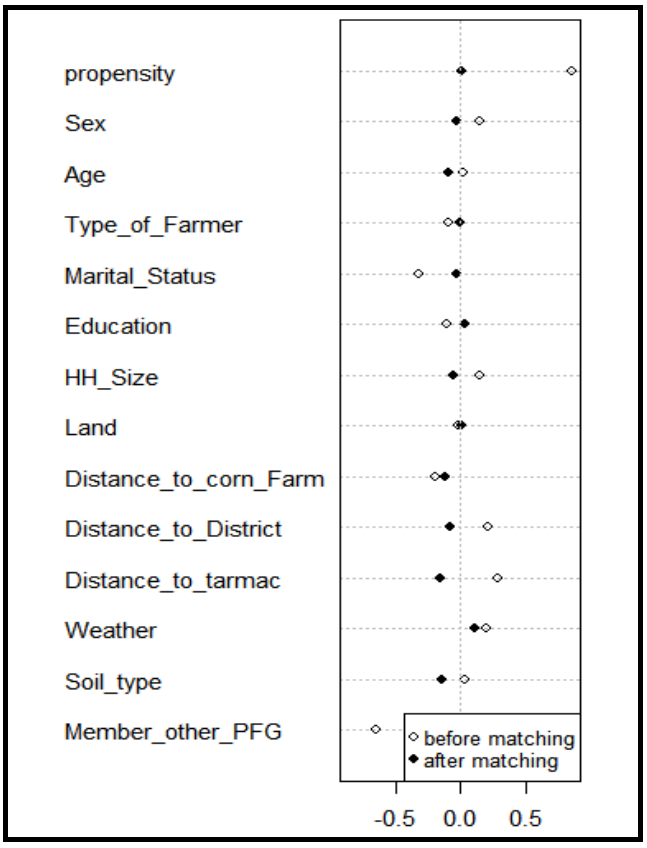


Figure 4. Dot plot of Standardized Mean Differences

In Figure 5, the standardized mean difference before and after matching are presented. The bolded lines are standardized differences that increase after matching.

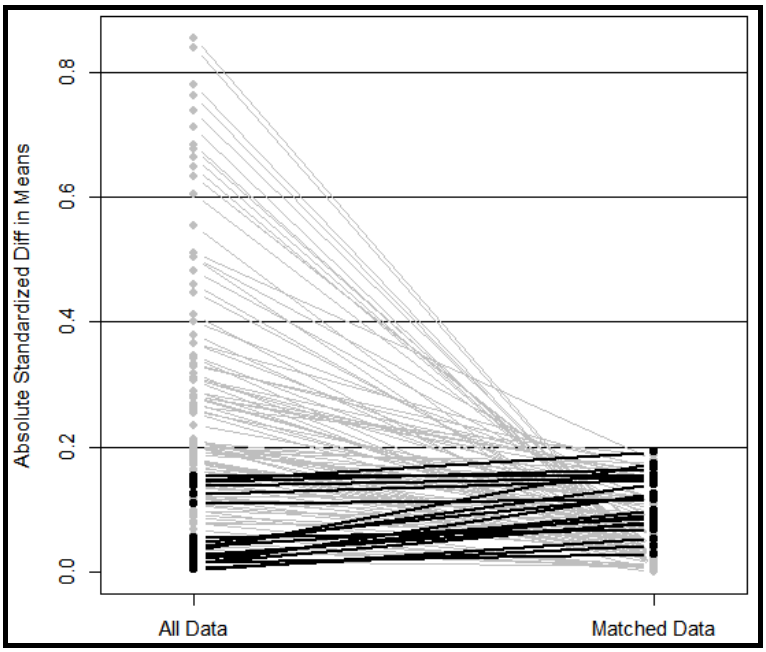


Figure 5. Line plot of Standardized Differences before and after Matching

All the univariate and multivariate indicate that there are no imbalances after matching. Out of 878, the matched samples were 252 for both control and treated farmers. From the control group, about 265 were unmatched and 94 for treated farmers were unmatched. About 15 samples (2 control and 13 treated) were discarded because they were outside the common support.

3.2 Effects of Intervention

The t test was adopted so as to compare the means of two groups (control and treatment). The test is used when the subjects are randomly assigned to two groups. Although the subjects were not randomly assigned into two groups, the adoption of PSM to some extent controls selection bias and make the difference between the groups to be due to treatment and not to other factors. The test results of the four items are presented in Table 2.

Table 2. Comparison of welfare of farmers

Welfare measure	***t***	***df***	***p-value***
Earnings from corn production	-1.697	460	0.090
Value of livestock owned	0.824	460	0.410
Value of household assets owned	0.366	460	0.715
Value of farm assets owned	-1.484	460	0.139

None of the four factors had significant result as the *p values* are greater than 0.05. This implies that the earning between farmers participating in DASIP are not significant different from those who do not participate in the programme. The same applies to the other three welfare measures. Looking at the sign of t values, it was found that farmers who do not participate in DASIP had high earnings from corn production compared to those participating in the programme. Also, the values of farm assets of non-participating farmers are higher than those of participant farmers. The participant farmers had higher values of livestock and household assets compared to non-participant farmers.

The findings show that DASIP has not contributed much to the welfare of farmers. This result is contrary to that obtained by Okorley et al. (2004) when studied the effects of FFS in Cocoa intervention. They find that capital assets of the farmers improved. Davis et al. (2010) find that the value of crops produced per acre, livestock value gain per capita, and agricultural income per capita increased significantly for FFS participants. The vast of the majority in literature find that FFS has improved the status of participants than non-participants. The nature and way the data were analysed is quite different to this paper. In literature, many authors focused on assessing knowledge and skills.

The results of this paper could be the results of most DASIP groups not to practice what they have acquired from DASIP. During the data collection, it was observed that groups have a tendency of changing group activities often. A group can cultivate corn this year but after just a year it can turn to be livestock keeper. There are limited groups which stick on the activities since the formation of the groups. The data were collected in rural area and average distance to district headquarter is 20 km.

The insignificant contribution of DASIP could be also contributed by unsatisfactory monitoring of the activities. Sometimes, the projects and activities which were performed by groups are not monitored and evaluated by the DASIP officials because of the scarcity of funds. Most of the farmers require maximum supervision in order to benefit from activities engaged. So, if there is a limited follow-up, there will be limited knowledge diffusion. Other factors such as different cultures, lack of commitment among group members, mistrust among group members, lack of water and land degradation affect both productivity and income of the farmers. A combination of all factors affects the outcome based of the four items.

4. Conclusion

Basing on the findings, the welfare of the farmers participating in DASIP is almost the same as that for those who do not participate in DASIP. These findings suggest that, still there is much to do with agriculture intervention in order to increase productivity. The study recommends that agriculture intervention should include a component of irrigation as one of the objectives. Also, activities performed by a group should last longer rather than changing the activities from time to time. Furthermore, there should be close monitoring and evaluation of group activities, sensitization and awareness of intervention should be advocated.

Acknowledgement

I would like to thank all my students who participated in the data collection exercise. These are Martin Dominic, Amos Justine, Sikujua Abdallah, Kway Yudathadei, Ramadhan Malimba, Masala Yohana, John Mwita, Shaban Juma and Amandus Nkenzidyo.

References

Braun, A., Jiggins, J., Röling, N., Berg, H. and Snijders, P. (2006). A Global Survey and Review of Farmer Field School Experiences. [Online] Available: http://www.ifad.org/rpr2011/background/1.pdf (August 24, 2011).

Bryson, A., Dorsett, R., & Purdon, S. (2002). The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies. Crown.

Davis, K., Nkonya, E., Kato, E., Mekonnen, D. A., Odendo, M., Miiro, R., & Nkuba, J. (2010). Impact of farmer field schools on agricultural productivity and poverty in East Africa. International Food Policy Research Institute (IFPRI).

Dinpanah, G., Mirdamadi, M., Badragheh, A., Sinaki, J. M., & Aboeye, F. (2010). Analysis of Effect of Farmer Field School Approach on Adoption of Biological Control on Rice Producer' Producer' Characteristics in Iran. *American-Eurasian J. Agric. & Environ. Sci., 7* (3), 247-254.

Duflo, E., & Kremer, M. (2003). Use of randomization in the evaluation of development effectiveness. Aper Prepared for the World Bank Operations Evaluation Development (OED) Conference on Evaluation and Development Effectiveness in Washington, DC.

Erickson, R. (2003). Review and Evaluation Technical Assistance No. 3383-PAK: Integrated

Pest Management. Asian Development Bank. Philippines.

Feder, G., Murgai, R., & Quizon, J. (2003). Sending farmers back to school: The impact of farmer field schools in Indonesia. *Rev. Agric. Econ., 26,* 1-18. http://dx.doi.org/10.1596/1813-9450-3022.

Godtland, E. M., Sadoulet, E., Janvry, A. D., & Murgai, R. (2004), The Impact of Farmer Field Schools on Knowledge and Productivity: A study of Potato Farmers in the Perivian Andes, The Univesrty of Chicago.

Heinrich, C. Maffioli, A., & Vázquez, G. (2010). A Primer for Applying Propensity-Score Matching: Impact-Evaluation Guidelines. Inter-American Development Bank.

LaLonde, R. J. (1986), Evaluating the Econometric Evaluations of Training Programs with Experimental Data, *The American Economic Review*, *76*(4), 604-620.

Okorley, E. L., Adjargo G., & Bosompem, M. (2004). The Potential of Farmer Field School in Cocoa Extension Delivery: A Ghanaian Case Study. *Journal of International Agricultural and Extension Education*, Volume 21, Issue 2. http://doi: 10.5191/jiaee.2014.20203.

Ooi, P. A., & Kenmore, P. E. (2005). Impact of Educating Farmers about Biological Control in Farmer Field Schools. Second International Symposium on Biological Control of Arthropods.

Quizon, J., Feder, G., & Murgai, R. (2001). Fiscal Sustainability of Agricultural Extension: The Case of the Farmer Field School Approach. *Journal of International Agricultural and Extension Education 8*(1), 13-24. http://citeseerx.ist.psu.edu/viewdoc/summary ?doi=10.1.1.473.3519.

Rola, A. C., S. B. Jamias, & J. B. Quizon (2002). Do farmer field school graduates retain and share what they learn: an investigation in Iloilo, Philippines. *Journal of International Agricultural and Extension Education* 9: http://doi: 10.5191/jiaee.2002.09108.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's Critique of Non-experimental estimators? *Journal of Econometrics, 125,* 305-353. http://dx.doi.org/10.2139/ssrn.286297.

Thoemmes, F. (2012). Propensity score matching in SPSS. [Online] Available: http://sourceforge.net/projects/psmspss/files/ (September 12, 2014).

Tripp, R., Wijeratne, M., & Piyadasa, V. H. (2005). What should we expect from farmer field schools? A Sri Lanka case study. *World Development - Volume* 33, no. 10, pp. 1705-1720. http://dx.doi.org/10.1016/j.worlddev.2005.04.012

Van den Berg, H. (2004). IPM Farmer Field Schools: A synthesis of 25 impact evaluations. Wageningen University, Report prepared for the FAO Global IPM Facility, Rome.

Waddington, H., Snilstveit, B., White, H., & Anderson, J. (2010). The Impact of Agricultural Extension Services. *International Initiative for Impact Evaluation*. http://dx.doi.org/10.4102/aej.v1i1.19.

Yeshwanth K. N. L. G. (2008). A Study on Knowledge and Adoption of Integrated Crop Management (ICM) Practices by the Participants of Farmers Field School (FFS) in Bellary District. Masters Dissertation, University of Agricultural Sciences, Dharwad.

Yamane, Taro. (1967). Statistics: An Introductory Analysis, 2nd Edition, New York: Harper and Row.

The American Journal of Surgery 219 (2020) 504–507

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

Bariatric surgery is associated with reduction in non-alcoholic steatohepatitis and hepatocellular carcinoma: A propensity matched analysis

Minyoung Kwak, J. Hunter Mehaffey, Robert B. Hawkins, Angel Hsu, Bruce Schirmer, Peter T. Hallowell*

University of Virginia Health System, Charlottesville, VA, USA

ARTICLE INFO

Article history:
Received 20 June 2019
Received in revised form
2 September 2019
Accepted 11 September 2019

Keywords:
Bariatric surgery
NASH
Non-alcoholic steatohepatitis
HCC
Hepatocellular carcinoma
Weight loss

ABSTRACT

Introduction: Obesity is a risk factor for non-alcoholic steatohepatitis (NASH) and hepatocellular carcinoma (HCC). Bariatric surgery can provide durable weight-loss, but little is known about the later development of NASH and HCC after surgery.

Methods: Bariatric surgery (n = 3,410) and obese controls (n = 46,873) from an institutional data repository were propensity score matched 1:1 by demographics, comorbidities, BMI, and socioeconomic factors. Comparisons were made through paired univariate analysis and conditional logistic regression.

Results: Total of 4,112 patients were well matched with no significant baseline differences except initial BMI (49.0 vs 48.2, $p = 0.04$). Bariatric group demonstrated fewer new-onset NASH (6 0.0% vs 10.3%, $p < 0.0001$) and HCC (0.05% vs 0.34%, $p = 0.03$) over a median follow-up of 7.1 years. After risk-adjustment, bariatric surgery was independently associated with reduced development of NASH (OR 0.52, $p < 0.0001$).

Conclusions: Bariatric surgery is associated with reduced incidence of NASH and HCC in this large propensity matched cohort. This further supports the use of bariatric surgery for morbidly obese patients to ameliorate NASH cirrhosis and development of HCC.

Introduction

Incidences of non-alcoholic steatohepatitis (NASH) and hepatocellular carcinoma (HCC) are increasing throughout the United States.[1,2] A large contributing factor may be the rise of obese adults, as this trend is increasingly affecting adolescents, as well. The progression of NASH from non-alcoholic fatty liver disease (NAFLD) occurs in approximately 10–25% of patients[3] and can lead to significant risks in liver-related mortality due to the development of hepatic fibrosis, cirrhosis, and hepatocellular carcinoma (HCC). It is projected that more than 10% weight loss is needed in order to improve NASH,[4] however, weight loss modification through lifestyle changes alone account for only 3–5% total body weight loss on average and does not provide durable weight loss over time.[5] Two first-line medications (Vitamin E and pioglitazone) have been used to augment this effect,[6] however there are concerns due their association with other cancers and morality risk, also in their lack of improving hepatic fibrosis.[7] Additionally, the effectiveness of these medications was only studied in non-diabetic patients which leaves limited options for the greater proportion of obese patients that are also diabetic.

Bariatric surgery has shown to provide sustained weight loss throughout the course of a patient's lifetime,[8] and most patients who are candidates for bariatric surgery have some degree of NAFLD.[6] Previous studies have shown that bariatric surgery not only improves steatosis in NASH, but may also improve hepatic fibrosis even in patients who may have other metabolic diseases including diabetes mellitus type II (DM2).[5,7,9,10] However, this was not a consistent finding since a few studies also showed worsening hepatic fibrosis over time.[5,7,9–11] It is due to this concern that despite guidelines suggesting the benefit of bariatric surgery in reducing the progression to NASH, there is still no definitive recommendation on its routine use.[6,11] This may, in part, have contributed to the overall decrease in the number of bariatric

* Corresponding author.
E-mail address: PTH2F@hscmail.mcc.virginia.edu (P.T. Hallowell).

https://doi.org/10.1016/j.amjsurg.2019.09.006
0002-9610/

surgery procedures performed in patients with NAFLD from 2004-2012.[12] Clearly more evidence is needed to further understand this association.

There is a lack of case-controlled studies in the literature highlighting the effect of bariatric surgery on NASH[5,7] and the few studies on bariatric surgery and HCC only analyzed its prevalence in a bariatric population and did not address the development of new HCC after bariatric surgery.[13,14] Most studies also have relatively short follow up times as little as one month to only five years.[5,14] We hypothesized that performing case-matching through propensity score analysis of all patients who received bariatric surgery with long-term follow up will be able to better elucidate the risk of developing NASH and HCC after bariatric surgery.

Materials and methods

Patients

All adult patients who underwent bariatric surgery (including RYGB, sleeve gastrectomy, and adjustable gastric banding) for morbid obesity ($n = 3,410$) at a single academic institution between 1985 and 2015 were identified retrospectively from a prospectively maintained database.[15] To identify an appropriate control group, an institutional clinical data repository (CDR) of all routine outpatient visits from the same academic institution was queried to identify a non-operative cohort of 46,873 morbidly obese patients who did not undergo bariatric surgery. Propensity-matched groups were then generated to facilitate adjusted comparisons between the operative and non-operative cohorts. The Institutional Review Board of the University of Virginia approved this study (#17132) with waiver of consent due to its retrospective nature.

Data collection

Patient demographics, BMI, relevant comorbidities (diabetes mellitus, hypertension, gastroesophageal reflux disease, congestive heart failure/coronary artery disease, current tobacco and alcohol use), and insurance status were captured through the CDR for all patients. Patients were excluded if they were less than 18 years old, prisoners, or had incomplete medical records. These baseline characteristics were collected at the time of initial diagnosis of morbid obesity (BMI $>40\ kg/m^2$) that were entered into the electronic medical record for non-operative control patients, and at the time of the preoperative appointment for the patients who received bariatric surgery. Diagnoses of NASH and HCC were identified by ICD codes through the CDR for both operative and non-operative patients before and after the time of bariatric surgery or diagnosis of obesity in the control group. New diagnoses of NASH or HCC were captured even if patients were treated at outside institutions through the inclusion of ICD codes upon follow up visits. For all patients with a diagnosis of HCC, tumor characteristics were collected, as available, via review of the electronic medical record including pathology reports, laboratory results, and radiologic imaging even from outside institutions, if applicable.

Statistical analyses

Patients were matched 1:1 with all model variables chosen a priori, including demographics (age, initial BMI, race), history of NASH or HCC, relevant comorbidities (DM2, hypertension, gastroesophageal reflux disease (GERD), congestive heart failure/coronary artery disease), relevant preoperative substance use (alcohol/tobacco), and time of follow up. Adequacy of the match was assessed by balance metrics, including standardized mean difference and histograms of propensity scores.[16]

The primary outcome of interest was the overall incidence of NASH or HCC between the operative and non-operative groups. Secondary outcomes included differences in tumor characteristics among patients diagnosed with HCC. Complete tumor characteristics were not available for all patients. Univariate analyses were performed using Chi-square or Fisher's exact tests for categorical variables and Wilcoxon rank-sum test for continuous variables to assess for statistical differences in demographics, outcomes, and tumor characteristics between the operative and non-operative patients. Within the propensity matched cohort, multivariate logistic regression was used to assess the association between bariatric surgery and the incidence of NASH. Variables in the model were selected a priori based on clinical risk factors for NASH, and performance of the model was assessed by calculating area under the curve. Statistical significance was defined with the standard, two-sided alpha value of <0.05. Statistical analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC).

Results

A total of 3,410 patients who received bariatric surgery were evaluated and compared with 45,750 obese control patients who did not receive bariatric surgery from the same institutional data repository. Both patient groups varied in all demographic factors, and in almost all baseline comorbidities evaluated (Supplemental Table 1).

Propensity case-matching resulted in the inclusion of 2,057 bariatric surgery patients and 2,055 control patients (Table 1). The two groups were well matched in all baseline characteristics except for initial BMI (Table 1, Supplemental Figure 1) with histograms of matched propensity scores shown in Supplemental Figure 2. The clinical implications for a median difference in BMI of $0.6\ kg/m^2$ is likely minimal. Both BMIs (47 vs $46.4\ kg/m^2$) would still be within the same extreme obesity class III designation. After propensity score matching, there were no differences in the baseline prevalence of NASH (279 vs 256, $p = 0.29$), DM2 (432 vs 460, $p = 0.28$), viral hepatitis (102 vs 104, $p = 0.88$), GERD (958 vs 971, $p = 0.66$), and alcohol use (34 vs 39, $p = 0.55$) between the bariatric surgery group and control group.

Of the 2,057 patients in the matched bariatric surgery group, 121 (5.9%) received sleeve gastrectomy, 1,617 (79%) received RYBG, 275 (13%) received laparoscopic gastric banding, and 44 (2%) received other bariatric procedures. The median follow-up time was 7.1 years after bariatric surgery or after the initial obese diagnosis in the control group.

Patients in the bariatric surgery group developed lower incidences of NASH (123 (6%) vs 212 (10%), $p < 0.0001$) compared to the propensity-matched control group. This was highly significant even when both groups were matched on demographics, related comorbidities, tobacco use, and the initial BMI just prior to surgery or with the obese diagnosis in the control group as highlighted in Table 1. Lower incidences of NASH were present in the bariatric surgery group compared to the matched control group during the duration of follow up as shown in Fig. 1. Bariatric surgery patients also progressed to decreased incidences of HCC (1 (0.05%) vs 7 (0.3%), $p = 0.03$) as only one patient was found to have HCC in the matched bariatric surgery group. Further statistical analysis on the differences of tumor characteristics were not performed due to the low number of cases of HCC in both groups.

After risk adjustment using conditional logistic regression which included clinical variables associated with obesity and NASH, bariatric surgery was found to be independently and highly associated with a decreased incidence of NASH by 48% (OR 0.52, 95% CI 0.40–0.68, $p < 0.0001$, Table 2). The same analysis also found viral hepatitis to have a negatively association with NASH (OR 0.23, 95%

Table 1
Comparison of baseline characteristics of the propensity-matched bariatric surgical and non-surgical groups.

	No Bariatric Surgery n = 2,055	Bariatric Surgery n = 2,057	*p* value
Age (years)	43 (21)	42 (15)	0.83
Female	1738 (85)	1709 (83)	0.19
White	1764 (86)	1780 (86)	0.52
Initial BMI (kg/m^2)	46.4 (16)	47 (9)	0.04
Government insurance	835 (41)	790 (38)	0.14
DM Type II	460 (22)	432 (21)	0.28
Viral Hepatitis	104 (5)	102 (5)	0.88
NASH	256 (12)	279 (14)	0.29
Hypertension	782 (38)	790 (38)	0.82
GERD	971 (47)	958 (47)	0.66
Current alcohol use	39 (2)	34 (2)	0.55
Current smoker	556 (27)	550 (27)	0.82
CHF/CAD	91 (4)	97 (5)	0.66
COPD	59 (3)	52 (2)	0.5
OSA	129 (6)	137 (7)	0.62
Degenerative Joint Disorder	376 (18)	413 (20)	0.15

Categorical variables listed as N (%) and continuous variables listed as median (IQR).
BMI = Body mass index; DM Type II = Diabetes mellitus type II; NASH = Non-alcoholic steatohepatitis; GERD = Gastroesophageal reflux disease; CHF = Congestive heart failure; CAD = Coronary artery disease; COPD = Chronic obstructive pulmonary disease; OSA = Obstructive sleep apnea; HCC = Hepatocellular carcinoma.

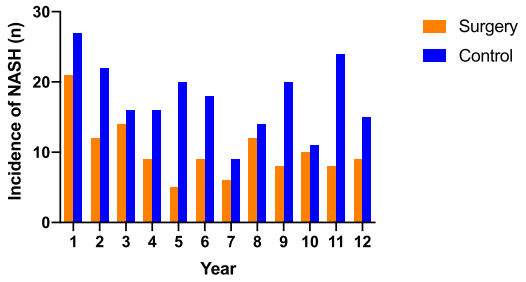


Fig. 1. Distribution of the Incidence of NASH in the Bariatric Surgery Group Compared to the Non-surgical Control Group Over Time.
Number of new NASH cases (n). Year is number of follow up years after bariatric surgery or obese diagnosis in the control group. NASH = Non-alcoholic steatohepatitis.

Table 2
Conditional logistic regression for risk of developing new non-alcoholic steatohepatitis after bariatric surgery (NASH). DM2 = Diabetes mellitus Type II; GERD = Gastroesophageal reflux disease.

Risk factors	Odds Ratio	95% CI		*p* value
Bariatric Surgery	0.52	0.40	0.68	<0.0001
Viral Hepatitis	0.23	0.12	0.48	<0.0001
DM Type 2	0.70	0.45	1.08	0.11
Female	0.70	0.40	1.2	0.21
White	1.55	0.85	2.82	0.15
Federal Insurance	1.35	0.92	1.98	0.13
Hypertension	0.85	0.54	1.33	0.47
Cardiac Disease	0.35	0.13	0.91	0.03
GERD	0.48	0.33	0.72	0.0003
Tobacco	1.28	0.85	1.95	0.24
Alcohol	0.85	0.31	2.38	0.76

CI 0.12–0.48, $p < 0.0001$) that was highly statistically significant. All viral subtypes of hepatitis were included in the analysis, and so it is unclear how many hepatitis patients had the Hepatitis C subtype compared to other subtypes that are less associated with NASH.

Discussion

Bariatric surgery was associated with fewer cases of NASH by 48% through risk-adjusted analysis compared to propensity score matched controls in this large cohort of patients with extended follow up. Through case-controlled propensity score matching, we found bariatric patients had decreased incidences of both NASH and HCC even when cases were matched on demographics and comorbidity risk factors such as DM2, GERD, alcohol and tobacco use. The one bariatric patient with HCC had a questionable diagnosis suggestive of HCC, but based on final histopathology, had other possible etiologies that also included cholangiocarcinoma and distant metastasis. Therefore, it is possible that the bariatric surgery group ultimately had no new cases of HCC during follow up. This aligns with a recent nationwide database study that showed decreased prevalence rates of HCC in bariatric surgery patients who were also propensity-matched with obese controls.[13]

Currently, NASH is the second leading cause of HCC requiring liver transplantation.[17,18] The increasing prevalence of NASH is predicted to overcome viral hepatitis as the leading cause of HCC since the advent of antiviral therapies have become more curative.[6] Our results showed bariatric surgery patients had a decreased probability of new NASH diagnoses after surgery compared to obese controls. Other meta-analyses involving liver pathology of post-bariatric surgery patients also found similar results[5,7,9,10]. One large meta-analyses which included 32 studies found bariatric surgery led to complete resolution of hepatic steatosis, inflammation, and balloon degeneration in the majority of patients, with significant decreases in NASH compared to baseline.[9] Another meta-analysis also found an overall decrease in the incidences of NASH after subgroup analysis even with different types of bariatric surgery procedures.[10] Unfortunately, we were not able to make a similar analysis due to the distribution of NASH and HCC among the different bariatric procedures in our database. Despite multiple studies showing overall positive findings in a majority of patients, a small group of patients also developed new or worsening hepatic fibrosis.[5,7,9–11,19] The cause for this discrepancy is still not well understood and may not be accounted for the different types of bariatric surgery alone. The risk of worsening or developing hepatic fibrosis is important to the safety and efficacy of bariatric surgery on patients with NASH cirrhosis. Previous studies show increased mortality risk after bariatric surgery in patients with NASH cirrhosis[20,21]. This largely contributes to why bariatric surgery is still not an established recommendation for NAFLD or NASH[6,11]. Future studies are needed to establish risk stratification for patients with NAFLD or NASH prior to bariatric surgery in order to identify patients more or less likely to develop and progress to NASH cirrhosis after surgery.

There are multiple theories regarding the pathogenesis of liver disease in obese patients that are related to metabolic syndrome

and insulin resistance. The increased release of free fatty acids and diacylglycerol from adipocytes, as well the secretion of pro-inflammatory cytokines such as TNF-alpha, IL-6, and leptin, can lead to a chronic inflammatory state that stimulate hepatic cell proliferation while also inhibiting apoptosis.[22] Higher leptin levels have also been implicated in the angiogenesis of HCC.[22] Since the risk factors of both HCC and NASH include obesity, it is unclear if the long-term weight loss resulting from bariatric surgery, or the weight loss itself, is associated with this decreased risk.

This study is limited by the retrospective design; however, the use of propensity matching accounted for important baseline characteristics in both groups, including initial BMI and comorbidities common in patients with morbid obesity. Unfortunately, we are unable to provide information regarding the matched interval weight loss of both the surgical and non-surgical groups for the duration of follow up since this was not consistently and accurately recorded in the retrospective review. Also, attempting to accurately report BMI at the time of new NASH or HCC diagnosis is highly susceptible to error, particularly in patients treated at outside institutions. In addition, we were surprised to find viral hepatitis was also a highly significant risk factor in our logistic regression analysis. It may be the thorough preoperative evaluation prior to bariatric surgery captured greater numbers of patient with baseline viral hepatitis compared to the control patients, and so had decreased incidences of new cases during the extended follow up period.

If bariatric surgery can improve the progression of NASH while also decreasing incidences of NASH and even possibly HCC, then there is no need to avoid bariatric surgery in patients with morbid obesity and NAFLD. We would not only be prohibiting the definitive weight loss option for these patients, but also prohibiting the harmful, obesity-related outcomes which would ultimately result in decreased quality of life and higher overall costs of their care.

Conclusion

Propensity match analysis of a large cohort of bariatric surgery patients compared with obese non surgery controls revealed patients who had undergone bariatric surgery had fewer new cases of NASH and HCC during with extended follow up. Further risk adjustment also showed bariatric surgery was associated with fewer cases of NASH by 48%. These results highlight the importance of bariatric surgery offering more than a procedure for sustained weight loss, but also in its potential to further abate obesity-related comorbidities, as well.

Conflicts of interest

The authors report no proprietary or commercial interest in any product mentioned or concept discussed in this article.

Acknowledgments

Funding Sources for Minyoung Kwak: NCI Cancer Center Support Grant P30 CA44579 Farrow Fellowship, University of Virginia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amjsurg.2019.09.006.

References

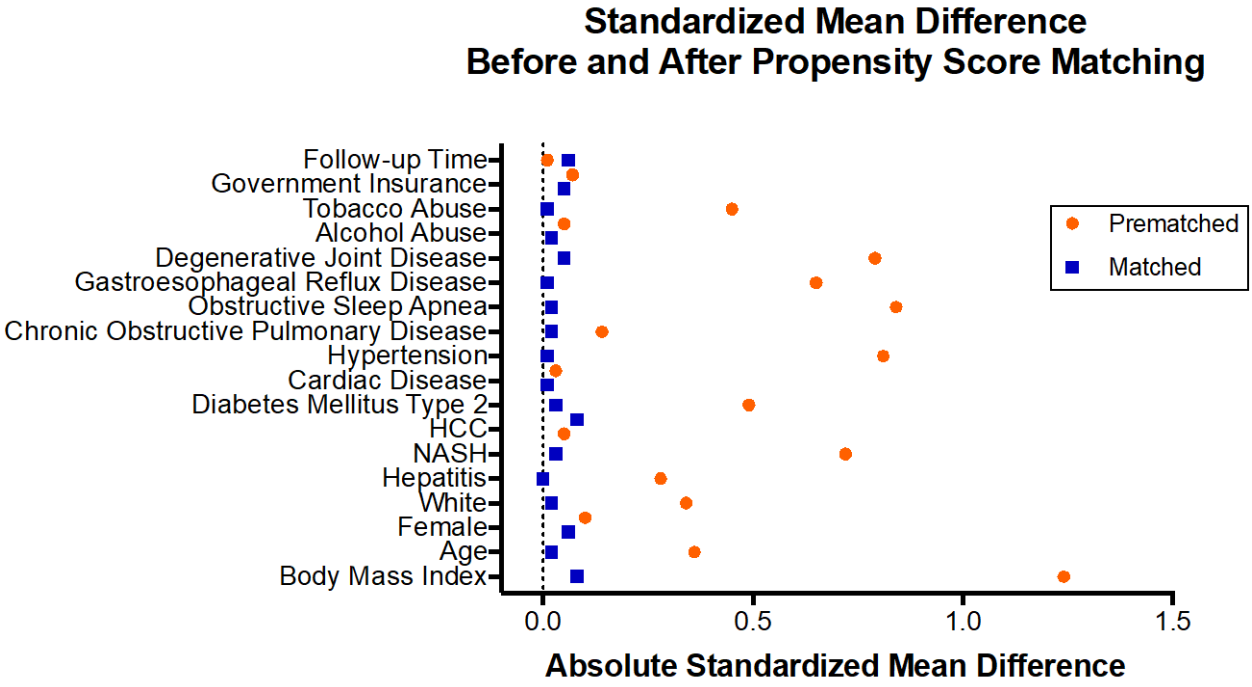
1. Mak L-Y, Cruz-Ramón V, Chinchilla-López P, et al. Global epidemiology, prevention, and management of hepatocellular carcinoma. In: *American Society of Clinical Oncology Educational Book*. vol. 38. 2018:262–279. https://doi.org/10.1200/EDBK_200939.
2. Lazo M, Clark JM. The epidemiology of nonalcoholic fatty liver disease: a global perspective. *Semin Liver Dis*. 2008;28(4):339–350. https://doi.org/10.1055/s-0028-1091978.
3. Milić S, Lulić D, Štimac D. Non-alcoholic fatty liver disease and obesity: biochemical, metabolic and clinical presentations. *World J Gastroenterol*. 2014;20(28):9330–9337. https://doi.org/10.3748/wjg.v20.i28.9330.
4. Glass LM, Dickson RC, Anderson JC, et al. Total body weight loss of $\geq$ 10 % is associated with improved hepatic fibrosis in patients with nonalcoholic steatohepatitis. *Dig Dis Sci*. 2015;60(4):1024–1030. https://doi.org/10.1007/s10620-014-3380-3.
5. Popov VB, Lim JK. Treatment of nonalcoholic fatty liver disease: the role of medical, surgical, and endoscopic weight loss. *J Clin Transl Hepatol*. 2015;3(3):230–238. https://doi.org/10.14218/JCTH.2015.00019.
6. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2018;67(1):328–357. https://doi.org/10.1002/hep.29367.
7. Rabl C, Campos GM. The impact of bariatric surgery on nonalcoholic steatohepatitis. *Semin Liver Dis*. 2012;32(1):80–91. https://doi.org/10.1055/s-0032-1306428.
8. Buchwald H, Estok R, Fahrbach K, et al. Weight and type 2 diabetes after bariatric surgery: systematic review and meta-analysis. *Am J Med*. 2009;122(3):248–256. https://doi.org/10.1016/j.amjmed.2008.09.041. e5.
9. Lee Y, Doumouras AG, Yu J, et al. Complete resolution of nonalcoholic fatty liver disease after bariatric surgery: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. October 2018. https://doi.org/10.1016/j.cgh.2018.10.017.
10. Hafeez S, Ahmed MH. Bariatric surgery as potential treatment for nonalcoholic fatty liver disease: a future treatment by choice or by chance? *J Obes*. 2013;2013:839275. https://doi.org/10.1155/2013/839275.
11. Chavez-Tapia NC, Tellez-Avila FI, Barrientos-Gutierrez T, Mendez-Sanchez N, Lizardi-Cervera J, Uribe M. Bariatric surgery for non-alcoholic steatohepatitis in obese patients. *Cochrane Database Syst Rev*. 2010;1:CD007340. https://doi.org/10.1002/14651858.CD007340.pub2.
12. McCarty TR, Echouffo-Tcheugui JB, Lange A, Haque L, Njei B. Impact of bariatric surgery on outcomes of patients with nonalcoholic fatty liver disease: a nationwide inpatient sample analysis, 2004-2012. *Surg Obes Relat Dis*. 2018;14(1):74–80. https://doi.org/10.1016/j.soard.2017.09.511.
13. Njei B, McCarty TR, Sharma P, et al. Bariatric surgery and hepatocellular carcinoma: a propensity score-matched analysis. *Obes Surg*. August 2018. https://doi.org/10.1007/s11695-018-3431-5.
14. Yang B, Yang HP, Ward KK, Sahasrabuddhe VV, McGlynn KA. Bariatric surgery and liver cancer in a consortium of academic medical centers. *Obes Surg*. 2016;26(3):696–700. https://doi.org/10.1007/s11695-016-2051-1.
15. Mehaffey JH, LaPar DJ, Clement KC, et al. 10-Year Outcomes After Roux-en-Y Gastric Bypass. *Ann Surg*. 2016;264(1):121–126. https://doi.org/10.1097/SLA.0000000000001544.
16. McMurry TL, Hu Y, Blackstone EH, Kozower BD. Propensity scores: methods, considerations, and applications in the journal of thoracic and cardiovascular surgery. *J Thorac Cardiovasc Surg*. 2015;150(1):14–19. https://doi.org/10.1016/j.jtcvs.2015.03.057.
17. Wong RJ, Cheung R, Ahmed A. Nonalcoholic steatohepatitis is the most rapidly growing indication for liver transplantation in patients with hepatocellular carcinoma in the U.S. *Hepatology*. 2014;59(6):2188–2195. https://doi.org/10.1002/hep.26986.
18. Wong RJ, Aguilar M, Cheung R, et al. Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology*. 2015;148(3):547–555. https://doi.org/10.1053/j.gastro.2014.11.039.
19. Mathurin P, Hollebecque A, Arnalsteen L, et al. Prospective study of the long-term effects of bariatric surgery on liver injury in patients without advanced disease. *Gastroenterology*. 2009;137(2):532–540. https://doi.org/10.1053/j.gastro.2009.04.052.
20. Mosko JD, Nguyen GC. Increased perioperative mortality following bariatric surgery among patients with cirrhosis. *Clin Gastroenterol Hepatol*. 2011;9(10):897–901. https://doi.org/10.1016/j.cgh.2011.07.007.
21. Jan A, Narwaria M, Mahawar KK. A systematic review of bariatric surgery in patients with liver cirrhosis. *Obes Surg*. 2015;25(8):1518–1526. https://doi.org/10.1007/s11695-015-1727-2.
22. Siegel AB, Zhu AX. Metabolic syndrome and hepatocellular carcinoma: two growing epidemics with a potential link. *Cancer*. 2009;115(24):5651–5661. https://doi.org/10.1002/cncr.24687.

Supplemental Table 1. Baseline Characteristics Between All Bariatric Surgical Patients and Non-surgical Controls.

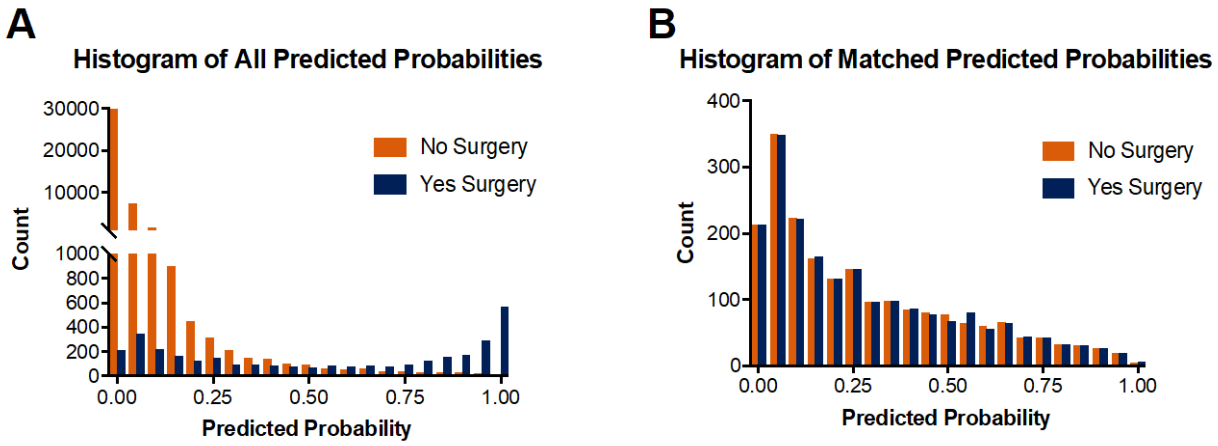
	No Bariatric Surgery n = 45,750	**Bariatric Surgery** n = 3,410	***p* value**
Age (years)	49 (23)	43 (16)	<0.0001
Female	30076 (66)	2768 (81)	<0.0001
White	34517 (75)	3008 (88)	<0.0001
Initial BMI (kg/m^2)	38.6 (7)	49 (12)	<0.0001
Government insurance	15961 (35)	1299 (38)	0.0002
DM Type II	4499 (10)	982 (29)	<0.0001
Hepatitis	1069 (2)	291 (8)	<0.0001
NASH	1003 (2)	867 (25)	<0.0001
Hypertension	7179 (16)	1741 (51)	<0.0001
GERD	11003 (24)	1846 (54)	<0.0001
Current alcohol use	1093 (2)	58 (2)	0.01
Current smoker	6008 (13)	1073 (31)	<0.0001
CHF/CAD	1834 (4)	158 (5)	0.07
COPD	611 (1)	116 (3)	<0.0001
OSA	234 (0.5)	934 (27)	<0.0001
Degenerative Joint Disorder	1652 (4)	1082 (32)	<0.0001
New NASH	3308 (7)	205 (6)	0.008
New HCC	210 (0.5)	12 (0.3)	0.37

Categorical variables listed as N (%) and continuous variables listed as median (IQR).
BMI = Body mass index; DM Type II = Diabetes mellitus type II; NASH = Non-alcoholic steatohepatitis; GERD = Gastroesophageal reflux disease; CHF = Congestive heart failure; CAD = Coronary artery disease; COPD = Chronic obstructive pulmonary disease; OSA = Obstructive sleep apnea; HCC = Hepatocellular carcinoma

Supplemental Figure 1: Balance assessment before and after propensity score matching showing the absolute standardized mean difference.



Supplemental Figure 2. Distribution of propensity scores **A.** before matching and **B.** after matching



Differences in sexually transmitted infection risk comparing preexposure prophylaxis users and propensity score matched historical controls in a clinic setting

Michalina A. Montaño[a], Julia C. Dombrowski[b,e], Sayan Dasgupta[c], Matthew R. Golden[a,b,e], Lisa E. Manhart[a,d], Lindley A. Barbee[b,e], Ann Duerr[a,c,d] and Christine M. Khosropour[a]

Objective: The aim of this study was to determine whether MSM using preexposure prophylaxis (PrEP) are at a higher risk of bacterial sexually transmitted infections (STIs) than MSM not using PrEP.

Design: Secondary analysis of longitudinal STI data obtained from MSM attending an STD Clinic in Seattle, Washington, USA, October 2011–September 2017.

Methods: We identified patients obtaining PrEP through the STD Clinic, and used propensity score matching to select a historical group of similar patients not using PrEP for comparison. We linked patient data with STI surveillance data to compare the incidence of chlamydia, gonorrhoea and early syphilis, and time to first symptomatic STI among PrEP users and nonusers.

Results: Three hundred and sixty-five PrEP users who picked up prescriptions and returned for follow-up and 730 propensity score matched nonusers were included in the analysis. Adjusted incidence rate ratios (aIRRs) for chlamydia, gonorrhoea and early syphilis were 3.2 [95% confidence interval (95% CI): 1.9–5.3], 2.8 (95% CI: 1.7–4.6) and 2.9 (95% CI: 1.5 – 5.6), respectively, comparing PrEP users to nonusers. Time to first symptomatic STI was shorter among PrEP users (120 days, 95% CI: 77 – 171) than among nonusers (185 days, 95% CI: 163–256).

Conclusion: Among MSM on PrEP, we observed a higher incidence of STIs and faster time to first symptomatic STI than MSM not using PrEP. PrEP may be a contributing factor in increasing STI rates among MSM.

AIDS 2019, **33**:1773–1780

Keywords: HIV, MSM, preexposure prophylaxis, sexually transmitted infections

Introduction

MSM are disproportionately impacted by HIV in the United States, accounting for 86% of new infections among men in 2016 [1]. Preexposure prophylaxis (PrEP) reduces the risk of HIV acquisition by up to 92% in MSM [2–4], is recommended by the Centers for Disease Control and Prevention (CDC) for HIV prevention among sexually active MSM [5] and is offered through a variety of clinical settings in the U.S. [6]. MSM are also

[a]Department of Epidemiology, [b]Department of Medicine, University of Washington, [c]Fred Hutchinson Cancer Research Center, [d]Department of Global Health, University of Washington, and [e]Public Health – Seattle and King County HIV/STD Program, Seattle, Washington, USA.

Correspondence to Michalina A. Montaño, Department of Epidemiology, University of Washington, 1959 NE Pacific Street, Box 357236, Seattle, WA 98195, USA.

Tel: +1 206 790 9576; fax: +1 206 543 8525; e-mail: micham@uw.edu
Received: 13 December 2018; revised: 3 April 2019; accepted: 12 April 2019.

DOI:10.1097/QAD.0000000000002281

ISSN 0269-9370

disproportionately impacted by bacterial sexually transmitted infections (STIs), the rates of which have increased substantially in the past decade. MSM accounted for 62% of gonorrhoea cases and 68% of primary and secondary syphilis cases among men in 2017 [7]. Between 2013 and 2017, the rate of primary and secondary syphilis among MSM increased by 64% [7].

Recent clinical data indicate that MSM using PrEP have an increased frequency of condomless anal sex after PrEP initiation [8–13], leading to concerns that PrEP may be contributing to increasing rates of bacterial STIs among MSM. But observational studies have reported mixed results regarding the association of PrEP use and increased rates of bacterial STIs among MSM [14–19]. However, these studies are limited by the lack of a comparison group of PrEP nonusers [16,17,19], failure to account for ascertainment bias resulting from increased STI screening among PrEP patients [15] or failure to adjust for secular STI trends [14,15,18]. Thus, the impact of PrEP use on STI risk remains unclear.

Since 2014, the Public Health – Seattle & King County (PHSKC) STD Clinic has provided PrEP to patients at a high risk of HIV. We used PHSKC STD Clinic data from PrEP users and a propensity score matched historical comparison group of nonusers to assess the impact of PrEP use on STI rates among MSM in the Seattle area. The primary objective of this study was to determine whether MSM using PrEP are at a higher risk of bacterial STIs than MSM not using PrEP.

Materials and methods

Study design, setting and population

This study was a secondary analysis of data from a cohort of MSM who initiated PrEP through the PHSKC STD Clinic between October 2014 and September 2017. As per PHSKC and Washington State (WA) PrEP guidelines [20], the PHSKC STD Clinic provides PrEP to patients at a high risk for acquiring HIV. This includes MSM and transgender persons who have sex with men who report any of the following risk factors in the past 12 months: diagnosis of rectal gonorrhoea or early syphilis; use of methamphetamine or amyl nitrites (poppers); or exchanging sex for money or drugs. PrEP is also recommended for patients with HIV-positive sexual partners who are not virally suppressed, and is provided to all interested African–American and Latino MSM. STD Clinic clinicians evaluate all patients for PrEP eligibility at routine visits, and those who meet the recommended criteria are offered PrEP through the clinic. Patients not meeting recommended criteria who are interested in initiating PrEP are referred to community PrEP providers. Throughout the study period (2011–2017), PHSKC has recommended that all MSM meeting the high-risk criteria defined above test for HIV and STIs quarterly, and has offered text-message reminders to encourage this [21].

PrEP visit procedures at our clinic have been described previously [9]. Briefly, PrEP patients are tested for HIV and STIs at their initial visit, and are given a 3-month prescription for PrEP. They return 1 month after PrEP initiation, and then quarterly for clinical follow-up and monitoring, including HIV and STI testing. We included PrEP patients (PrEP users) in this analysis if they were HIV-negative MSM who initiated PrEP through the STD Clinic between October 2014 and September 2017, who picked up their initial 3-month prescription, and returned for their first follow-up visit at 1 month. Our comparison group (PrEP nonusers) was composed of HIV-negative MSM who attended the STD Clinic between October 2011 and September 2014 (the period prior to PrEP availability at the clinic), propensity score matched to PrEP users. Our rationale for this approach was twofold: First, our clinic provides PrEP to patients at a high risk for HIV acquisition, and a comparison group of current clinic patients who are not on PrEP would be at a lower risk of HIV and STI acquisition than current PrEP users. Second, although PrEP is widely available in King County, our data only include PrEP status for STD Clinic patients. With no way to ascertain PrEP status for MSM outside of the STD Clinic population, we chose to compare our PrEP patients to propensity score matched historical controls from our clinic.

Propensity score matching and comparison group formation

We used propensity score matching to select a group of comparison patients who were most similar to our PrEP users, in order to identify a group of patients who would have likely been on PrEP in our STD Clinic if it had been available at the time they attended the clinic. The goal of propensity score matching is to approximate the effect of randomization by balancing observed covariates between study groups. We included 62 variables in our propensity score model, related to PrEP eligibility, sexual behaviour, STI risk, recent STI diagnoses, demographic characteristics and reasons for visiting the clinic (Supplementary Table, http://links.lww.com/QAD/B488). Data for the propensity score model were from the initial PrEP visit for PrEP users, and from the first STD clinic visit between October 2011 and September 2014 for PrEP nonusers. We used descending, nearest-neighbour matching without replacement to select two matched PrEP nonusers for each PrEP user based on their propensity scores [22,23]. After matching, we graphically compared the distribution of propensity scores in both groups (Supplementary Figure, http://links.lww.com/QAD/B489), and used graphs of standardized differences to check covariate balance between groups. Standardized differences of 10% or less were judged to be negligible and indicate good covariate balance [24,25]. Covariates with standardized

differences greater than 10% were added to final outcome models one at a time, and any that caused an absolute change in effect size of greater than 5% were included as covariates in final adjusted outcome models.

Data sources, measures and data linkage

For both PrEP users and PrEP nonusers, we obtained data on clinic attendance, visit dates, PrEP status and prescription fills from STD Clinic electronic patient records. Data on bacterial STI diagnoses were obtained from the Public Health Issue Management System (PHIMS), the electronic STI surveillance system used in WA. This system is separate from STD Clinic patient records, and only captures positive STI test results. WA laws require laboratories and medical providers to report all cases of chlamydia, gonorrhoea and syphilis to local health authorities who subsequently provide data to the WA Department of Health via PHIMS. The PHIMS data included all positive STI laboratory tests reported in King County during the time frame under study, including anatomic site of infection and diagnosis date. Thus, we used PHIMS to identify all STIs diagnosed during the study period, including those diagnosed outside the STD clinic. We defined early syphilis as primary, secondary or early latent stage syphilis. As most cases of urethral gonorrhoea and primary and secondary syphilis among men are symptomatic, we defined symptomatic STIs as cases of urethral gonorrhoea or primary or secondary syphilis [26,27]. We obtained the data included in the propensity score model from STD Clinic electronic medical records, standardized, comprehensive intake forms filled out by patients at all clinic visits, and from PHIMS.

We matched STD Clinic visit data to PHIMS STI case report data to identify incident STIs diagnosed during the study period. We linked these two data sources on first name, last name and date of birth. Matching was performed using fastLink, a probabilistic record linking package for R [28]. FastLink utilizes a Felligi–Sunter probabilistic record linkage model with an expectation-maximization algorithm. We used the default Jaro–Winkler method to measure agreement for partial matches, and set the lower bound for posterior match probability to 0.85.

Sexually transmitted infection testing

Quarterly STI tests are recommended for all clinic PrEP patients [29], and are performed as part of routine clinical care. Patients are screened for chlamydia and gonorrhoea at each anatomic site (urethra, pharynx, rectum) based on reported exposure. Urine samples, and urethral, pharyngeal and rectal swabs were tested for gonorrhoea and chlamydia using nucleic acid amplification testing (APTIMA Combo 2; Hologic, Inc, Marlborough, Massachusetts, USA). Blood samples were tested for syphilis using a quantitative rapid plasma reagin (Beckton Dickinson, Franklin Lakes, New Jersey, USA) test, with confirmatory tests performed using *Treponema pallidum* particle agglutination assay (Fujirebio Inc., Tokyo, Japan). All cases of syphilis in King County are staged by a disease intervention specialist based on laboratory and clinical findings.

Statistical analysis and follow-up time calculation

We used Poisson regression to compare incidence of bacterial STIs between PrEP users and PrEP nonusers. PrEP users were followed from their first prescription fill through 90 days after their last PrEP clinic appointment. Nonusers were followed for 1 year from their first clinic visit between October 2011 to September 2014. We limited the follow-up period for the comparison group to 1 year to minimize the likelihood that nonusers who moved out of King County would accrue follow-up time but not contribute STI diagnoses to incidence calculations. Any incident diagnoses of chlamydia, gonorrhoea or early syphilis were counted as outcomes with one exception: a diagnosis of an STI that occurred just after another diagnosis of the same STI (21 days for chlamydia, 14 days for gonorrhoea, 28 days for early syphilis) was not counted. Because this resulted in a window after each STI diagnosis wherein additional cases could not be counted, we subtracted the relevant number of days from the denominator for each diagnosis (e.g. subtracted 21 days from the denominator for each chlamydia diagnosis). We ran separate Poisson regression models for each STI, symptomatic STIs in combination and each anatomic site of infection for chlamydia and gonorrhoea. We clustered by patient ID to account for 43 PrEP users who were also included in the comparison group. We adjusted models for secular STI trends by including a continuous variable for annual incidence of each STI (or site-specific STI) in the general MSM population in King County [30]; these data were obtained from the PHSKC STD Epidemiologist. Rectal gonorrhoea and early syphilis models were not adjusted for secular trends due to model instability arising from years with zero cases among study patients.

To account for ascertainment bias resulting from more frequent STI screening among PrEP users, we compared time to first symptomatic STI between PrEP users and nonusers using Kaplan–Meier survival analysis. For this analysis, PrEP users were followed from their first prescription fill date and PrEP nonusers were followed from their first clinic visit date, with follow-up time for both groups censored at the first diagnosis of urethral gonorrhoea or primary or secondary syphilis. We used a log rank test to compare median time to first symptomatic STI between the groups. We compared the cumulative probability of experiencing a symptomatic STI between the groups during 1 year of follow-up using a Kaplan–Meier failure curve.

We used Stata version 15.1 (College Station, Texas, USA) for all analyses. Two-sided statistical tests were performed

at a significance level of 0.05. This study was approved by the University of Washington Institutional Review Board.

Results

Between October 2014 and September 2017, 557 MSM were prescribed PrEP through the PHSKC STD Clinic. Ninety-one patients were excluded from this analysis because they did not pick up their first prescription, and a further 101 did not return for their first follow-up visit, leaving 365 PrEP users in our analytic sample. Of these, 15 initiated PrEP in 2014, 108 in 2015, 121 in 2016 and 121 in 2017. For each PrEP user in our analysis, we selected two propensity score matched PrEP nonusers, resulting in a comparison group of 730 patients. Of these, 79 had their initial clinic visit in 2011, 262 in 2012, 223 in 2013 and 166 in 2014. Among PrEP users, the median length of time on PrEP was 292 days [interquartile range (IQR): 117–561]. The two study groups were balanced in terms of age, race/ethnicity, PrEP eligibility criteria and sexual behaviour (Table 1). The mean age in both groups was 30 years. Half of each group was white, non-Hispanic, and approximately 15% were diagnosed with syphilis in the past year. Men in both groups had a mean number of five male sexual partners in the prior 2 months and three-quarters of men reported receptive anal sex in the past year.

The c-statistic for our propensity score model was 0.81. The distribution of propensity scores was similar in both groups (Supplementary Figure, http://links.lww.com/QAD/B489). Fifty-two of the 62 covariates included in the propensity score were balanced between the two groups (Supplementary Table, http://links.lww.com/QAD/B488). All 10 unbalanced covariates resulted in absolute changes in point estimates of less than 5% after inclusion in adjusted models, so none were included in final outcome models.

Table 2 provides incidence and adjusted incidence rate ratio (aIRR) of each STI, comparing PrEP users to PrEP nonusers. Among PrEP users, incidence of chlamydia, gonorrhoea and early syphilis was 45.2, 37.1 and 6.9 per 100 person-years, respectively. PrEP users had an approximately three-fold higher incidence rate than PrEP nonusers for chlamydia (aIRR: 3.2; 95% CI: 1.9–5.3), gonorrhoea (aIRR: 2.8; 95% CI: 1.7–4.6) and early syphilis (aIRR: 2.9; 95% CI: 1.5–5.6). STI incidence among PrEP users was highest for rectal chlamydia (38.0

Table 1. Characteristics of preexposure prophylaxis users and propensity score matched historical comparison patients (preexposure prophylaxis nonusers) at the time of their initial Public Health – Seattle & King County STD clinic visit (*N* = 1095).

	PrEP users (*N* = 365)		PrEP nonusers (*N* = 730)		
Characteristic	*N*	%	*N*	%	*P*
Age (Mean, SD)	30.6	8.7	30.1	8.6	0.33
Race/Ethnicity[a]					
White, non-Hispanic	183	50.1	390	53.4	0.11
Black/African–American, non-Hispanic	25	6.8	43	5.9	
Hispanic	90	24.7	137	18.8	
Asian/Pacific Islander, non-Hispanic	41	11.2	76	10.4	
Mixed Race/Other, non-Hispanic[b]	14	2.7	42	4.7	
Unknown, non-Hispanic	12	3.3	42	5.8	
PrEP eligibility criteria[c]					
Rectal gonorrhoea diagnosis in the past year[d]	108	29.6	221	30.3	0.82
Early syphilis diagnosis in the past year[d]	58	15.9	106	14.5	0.55
Methamphetamine use in the past year	36	9.9	81	11.1	0.53
Popper use in the past year	175	47.9	313	42.9	0.11
Sex work in the past year[e]	11	3.0	21	2.9	0.90
Number of male sexual partners[f] (Mean, SD)	5.0	11.9	5.3	8.5	0.74
Sexual behaviour in past 12 months					
Receptive anal sex	272	74.5	565	77.4	0.29
With HIV-positive partner	46	12.6	81	11.1	0.46
With HIV-negative partner	233	63.8	502	68.8	0.10
With unknown status partner	91	24.9	169	23.2	0.51
Insertive anal sex	261	71.5	550	75.3	0.17
With HIV-positive partner	70	19.2	147	20.1	0.71
With HIV-negative partner	240	65.8	519	71.1	0.07
With unknown status partner	98	26.8	196	26.8	1.0

PHSKC, Public Health – Seattle & King County; PrEP, preexposure prophylaxis; SD, standard deviation.
[a]Categories for race/ethnicity are mutually exclusive.
[b]Other includes self-reported other race and Native American and Alaskan Native.
[c]PrEP eligibility in the PHSKC STD Clinic; categories are not mutually exclusive.
[d]Includes diagnosis at current visit.
[e]Sex work is defined as giving or receiving money or drugs in exchange for sex.
[f]Self reported, in prior 2 months.

Table 2. Incidence of bacterial sexually transmitted infections among preexposure prophylaxis users and nonusers attending the Public Health – Seattle & King County STD clinic (2011–2017), *N* = 1095.

STI	PrEP users (*N* = 365) Incidence, per 100 person-years	Nonusers (*N* = 730) Incidence, per 100 person-years	IRR	95% CI	aIRR[a]	95% CI
Chlamydia	45.2	14.4	3.1	2.4–4.2	3.2	1.9–5.3
Rectal	38.0	10.4	3.7	2.6–5.1	3.7	1.9–7.3
Urethral	7.4	3.7	2.0	1.1–3.6	2.2	1.1–4.5
Pharyngeal	3.8	2.5	1.6	0.8–3.2	1.6	0.7–3.8
Gonorrhoea	37.1	17.7	2.1	1.6–2.8	2.8	1.7–4.6
Rectal	20.7	9.8	2.1	1.5–3.0	2.1[b]	1.5–3.0
Urethral	9.9	6.1	1.6	1.0–2.6	1.5	0.6–4.1
Pharyngeal	16.2	9.2	1.8	1.2–2.5	2.0	1.1–3.5
Early syphilis[c]	6.9	2.3	2.9	1.5–5.6	2.9[b]	1.5–5.6
Symptomatic[d]	13.1	7.3	1.8	1.2–2.7	2.4	0.9–6.3

aIRR, adjusted incidence rate ratio; CI, confidence interval; IRR, incidence rate ratio; PrEP, preexposure prophylaxis.
[a]Models for incidence rate ratios are clustered by patient ID and adjusted for the annual incidence for each STI among all MSM in King County, for each year in 2011–2017 to account for secular STI trends, unless otherwise specified.
[b]Models adjusted for secular STI trends were unstable due to zero cases among study patients in some years. Models for rectal gonorrhoea and early syphilis are clustered by patient ID, but not adjusted for annual MSM incidence.
[c]Includes primary, secondary and early latent.
[d]Includes urethral gonorrhoea and primary and secondary syphilis.

per 100 person-years) and rectal gonorrhoea (20.7 per 100 person-years), which were both significantly higher than among PrEP nonusers. Incidence of urethral gonorrhoea, pharyngeal chlamydia and symptomatic STIs (urethral gonorrhoea and primary/secondary syphilis) was higher among PrEP users than nonusers, though not statistically significantly. Adjustment for secular trends in STI incidence had a minimal impact on IRRs.

The median time to first symptomatic STI among PrEP users was 120 days (95% CI: 77–171), compared with 185 days among PrEP nonusers (95% CI: 163–256; log-rank $P < 0.01$). The cumulative probability of experiencing a symptomatic STI by the end of 1 year of follow-up was nearly twice as high among PrEP users compared with PrEP nonusers (12 vs. 7%, respectively; Fig. 1).

Discussion

We found that, compared with MSM not using PrEP, PrEP users had a two to three-fold higher incidence of nearly all bacterial STIs, a higher incidence of symptomatic STIs, a 50% faster median time to first symptomatic STI and almost double the probability of experiencing a symptomatic STI during the first year of follow-up. These findings are consistent with recent data from our clinic showing increased frequency of

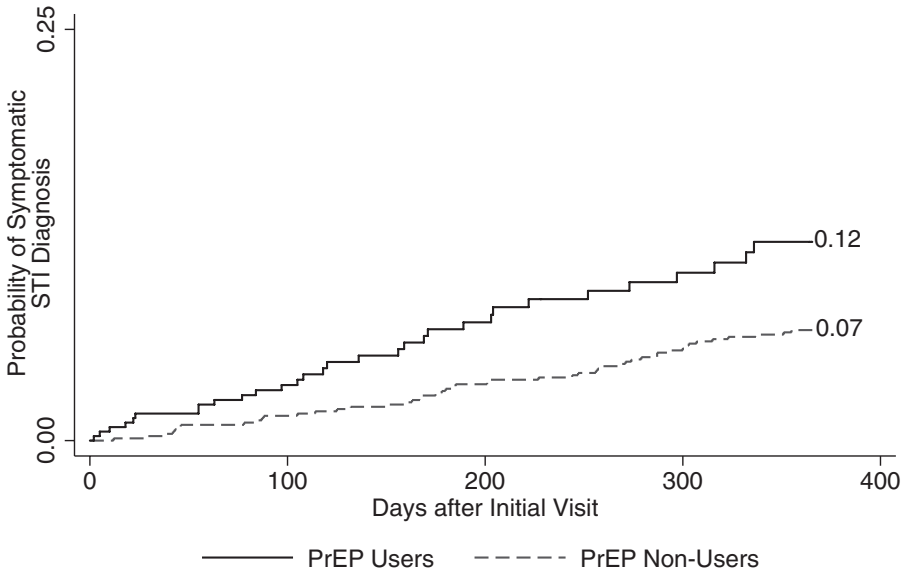


Fig. 1. Cumulative probability of symptomatic sexually transmitted infection diagnosis during the first year after initial clinic visit.

condomless sex among PrEP users [9], and support reports from prior studies of high STI incidence among PrEP users. Further, our findings suggest that PrEP use is associated with a higher risk of bacterial STIs independent of bias resulting from increased STI screening among PrEP users and secular trends.

Recent studies, including two meta-analyses of STI incidence among PrEP users, report incidence of bacterial STIs among PrEP users in the range of 38.0–56.7 per 100 person-years, 37.5–51.7 per 100 person-years and 9.1–14.5 per 100 person-years for chlamydia, gonorrhoea and syphilis, respectively [15–17,31]. Incidence of chlamydia and gonorrhoea among PrEP users in our study was similar, while incidence of early syphilis was slightly lower (45.2, 37.1 and 6.9 per 100 person-years, respectively). The majority of these studies did not include comparison groups of MSM not using PrEP [16,17,31]. However, the meta-analysis by Kojima *et al.* [15] compared pooled incidence estimates for MSM using PrEP and MSM not using PrEP from several studies of STI incidence among MSM.

In the meta-analysis by Kojima *et al.* [15], the pooled IRRs comparing STI incidence of PrEP users to nonusers, 11.2, 25.3 and 44.6 for chlamydia, gonorrhoea and syphilis, respectively [15], are substantially higher than those observed in our study. This is likely due to differences between our propensity score matched comparison group and the pooled comparison group of PrEP nonusers included in the meta-analysis. Most studies included in the meta-analysis did not include direct comparison of PrEP users and nonusers. Further, the studies included in the pooled incidence estimate for PrEP nonusers were extremely heterogeneous; studies included a wide range of behavioural criteria, some included HIV-positive men, and some dated back to 1998, when syphilis incidence was at its lowest point in the past 20 years. Secular trends in STI incidence, differences in screening frequency and differences in baseline sexual behaviour may have impacted the comparison of PrEP users and nonusers.

Two recent observational studies not included in the above meta-analysis used within-individual comparison to measure the impact of PrEP use on STI risk, comparing STI incidence during pre and post-PrEP initiation periods within the same group of participants [14,18]. Although their reported incidence during PrEP use is somewhat similar to our group of MSM using PrEP, Beymer *et al.* [14] observed decreases in incidence between the pre-PrEP and during-PrEP periods for all STIs except syphilis. However, their inclusion of STIs diagnosed at PrEP initiation in the pre-PrEP period may have contributed to their observation of decreases in incidence. We excluded STIs diagnosed at the initial visit from our incidence calculations. Nguyen *et al.* [18] similarly compared incidence of bacterial STIs among MSM using PrEP to the same group of MSM prior to PrEP use. In that study, only chlamydia was found to have higher incidence during PrEP use than before PrEP use (any site aIRR: 1.74; 95% CI: 1.02–2.96). In contrast, we observed a two to three-fold increase in incidence for most STIs included in our analyses.

Differences between our study findings and those mentioned above are likely a result of two key methodological differences. First, propensity score matching allowed us to identify a comparison group as similar as possible to our PrEP users with respect to recent STI diagnoses, sexual behaviour, HIV risk and other characteristics likely associated with PrEP use. This resulted in a group of PrEP nonusers who were most likely to have been using PrEP if it had been available to them. Second, of the three studies above that compared PrEP users with nonusers, only one addressed the issue of ascertainment bias resulting from higher STI screening frequency among PrEP users. Nguyen *et al.* [18] adjusted their models for the number of screening visits in the pre and post-PrEP initiation periods. Lacking data on frequency of STI testing among PrEP nonusers, we investigated the impact of PrEP use on risk of urethral gonorrhoea and primary and secondary syphilis, which are most frequently symptomatic and result in care-seeking, using survival analysis to measure time to first symptomatic STI among PrEP users and nonusers. We chose to focus on symptomatic STIs because detection of these infections would be unlikely to be affected by an increase in screening frequency. Indeed, our hypothesis that PrEP users would experience faster time to first symptomatic STI was supported by our findings, which provide evidence that PrEP use is associated with an increased risk of symptomatic STIs independent of increased screening frequency. These approaches allowed us to address both the difficulty of comparison group formation and the possibility of ascertainment bias, two limitations that have been present in many past studies of PrEP use and STI risk.

This study has a number of strengths. First, to our knowledge, this is the first study of the impact of PrEP on STI risk to compare time to first symptomatic STI between PrEP users and nonusers as a means of addressing ascertainment bias. Second, we adjusted for secular trends by including annual incidence of each STI among MSM in King County, which we were able to measure due to robust and detailed STI surveillance data from PHIMS. Third, our use of propensity score matching approximates the effect of randomization, and allowed us to balance a large number of observed characteristics between study groups. Further, this method of comparison group formation enabled us to identify PrEP nonusers who were most likely to have been on PrEP if it had been available. Finally, access to case report data from our state STI surveillance system allowed more complete ascertainment of STI outcomes by enabling us to identify STIs diagnosed outside of our clinic.

Our results should be interpreted in light of several limitations. First, observed incidence among PrEP users and comparison of incidence between PrEP users and nonusers is likely to be impacted by ascertainment bias resulting from higher frequency of STI testing among PrEP users. Because we did not have data available on STI screening frequency for either group, we attempted to address this by including an analysis of the incidence of symptomatic STIs, and comparison of time to first symptomatic STI. Second, propensity score matching is predicated on the assumption that the covariates included in the propensity score model can predict likelihood of PrEP use. If incorrect, this assumption may have resulted in biased estimates of the impact of PrEP on STI risk. Third, we were unable to adjust all statistical models for secular STI trends. However, the incidence in the two propensity-score matched groups does suggest that the magnitude of the relative association would have been similar to that which was reported in the fully adjusted models. Further, it is possible that the inclusion of yearly STI incidence among the general population of MSM failed to completely adjust for secular STI trends. However, insofar as PrEP played an important role in increasing STI incidence in MSM, it is also possible that our adjustment for secular trends represents an overadjustment. Finally, the generalizability of these results to other MSM populations outside of Seattle is uncertain.

In summary, we observed an increased STI risk among PrEP users relative to nonusers. Incidence of bacterial STIs among MSM has been increasing over the past decade, both in King County and nationally [7,30], and evidence that PrEP may be contributing to these increases is troubling. However, our findings do not negate the tremendous success of PrEP as a tool for HIV prevention. In light of centrality of PrEP as an HIV prevention tool, new and creative STI prevention efforts among PrEP users are needed. Our results highlight the importance of ongoing screening and treatment of STIs among PrEP users. Because these are components of PrEP patient care, it may be that PrEP programmes can be leveraged to continue to engage MSM in more comprehensive STI prevention programmes in the future. The success of PrEP programmes may provide a unique opportunity to design and implement novel interventions to address increasing STI rates in this population.

Acknowledgements

We wish to thank T. Avoundjian for his help with record linkage, and the staff at the PHSKC STD Clinic, including C. Thibault and J. Dimer for data management support, and C. Malinski, G. Afful, M. Barry, S. Herrmann, N. Ocbamichael and D. Spellman, for their work with patients in the PrEP clinic. This work was funded by Public Health – Seattle and King County; the National Institutes of Health (NIH) [F31 MH114892 trainee support to M.A.M.]; and the University of Washington Center for AIDS Research, an NIH-funded programme [P30 AI027757] which is supported by the following NIH Institutes and Centers: National Institute of Allergy and Infectious Diseases, National Cancer Institute, National Institutes of Mental Health, National Institute on Drug Abuse, National Institute of Child Health and Human Development, National Heart, Lung, and Blood Institute, and National Institute on Aging.

Author Contributions: M.A.M., C.M.K., J.C.D., and A.D. contributed to the conception of the study and analysis plan. C.M.K. and S.D. provided statistical consultation during analysis. M.A.M. merged datasets and performed the final analysis. M.A.M. and C.M.K. wrote and revised initial drafts of the manuscript. All authors contributed to study design and interpretation of results, reviewed the manuscript, provided edits, and approved the final paper.

Funding: C.M.K., M.R.C., L.A.B. and L.E.M. have received donations of specimen collection kits and reagents from Hologic, Inc. unrelated to this work. L.E.M. has received speaker's fees from Hologic, Inc. unrelated to this work. J.C.D. has conducted studies unrelated to this work funded by grants to the University of Washington from Hologic, Curatek and Quidel, and has received a speaker's honorarium and travel support for a meeting on retention in HIV care from Gilead. M.R.G. has received research support from GSK.

Conflicts of interest

All other authors declare that they have no conflict of interest.

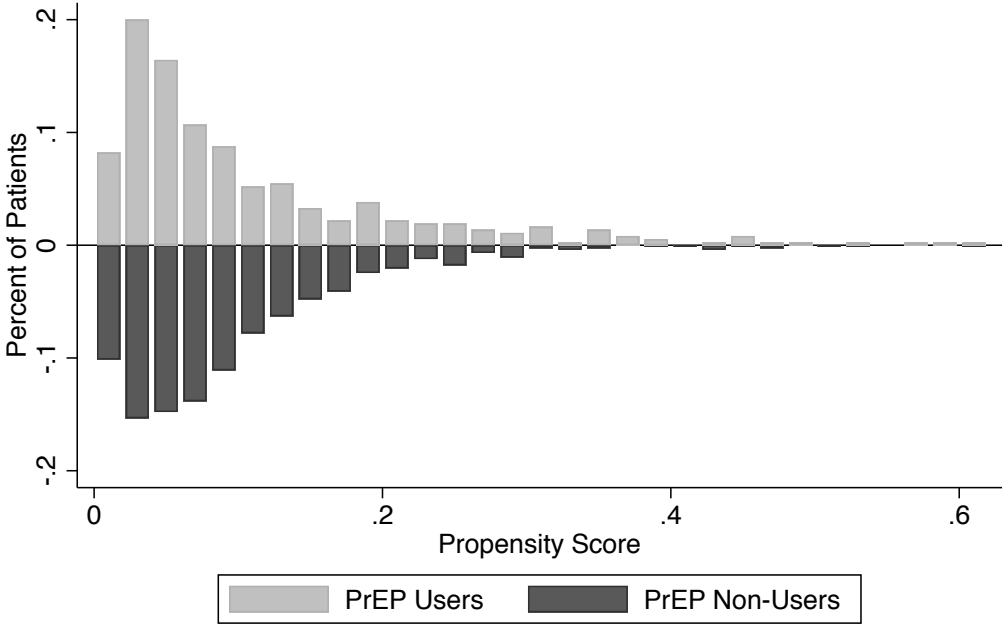
This work was presented in part at IUSTI 2018; Dublin, Ireland; 27–30 June 2018.

References

1. Centers for Disease Control and Prevention. HIV Surveillance Report, 2016; vol. 28. http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. Published November 2017. [Accessed November 2018].
2. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, *et al*. **Preexposure chemoprophylaxis for HIV prevention in men who have sex with men.** *N Engl J Med* 2010; **363**:2587–2599.
3. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, *et al*. **Uptake of preexposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study.** *Lancet Infect Dis* 2014; **14**:820–829.
4. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, *et al*. **Preexposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial.** *Lancet (London, England)* 2015; **387**:53–60.

5. Centers for Disease Control and Prevention: US Public Health Service: Preexposure prophylaxis for the prevention of HIV infection in the United States—2017 Update: a clinical practice guideline. https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2017.pdf. Published March 2018. [Accessed November 2018]
6. Marcus JL, Volk JE, Pinder J, Liu AY, Bacon O, Hare CB, *et al*. **Successful implementation of HIV preexposure prophylaxis: lessons learned from three clinical settings.** *Curr HIV/AIDS Rep* 2016; **13**:116–124.
7. Centers for Disease Control and Prevention. *Sexually transmitted disease surveillance 2017*. Atlanta: U.S. Department of Health and Human Services; 2018.
8. Hoornenborg E, Coyer L, Van Laarhoven A, Achterbergh R, De Vries H, Prins M, *et al*. **Change in sexual risk behaviour after 6 months of preexposure prophylaxis use: results from the Amsterdam preexposure prophylaxis demonstration project.** *AIDS* 2018; **32**:1527–1532.
9. Montaño MA, Dombrowski JC, Dasgupta S, Golden MR, Duerr A, Manhart LE, *et al*. **Changes in sexual behavior and STI diagnoses among MSM initiating PrEP in a clinic setting.** *AIDS Behav* 2019; **23**:548–555.
10. Newcomb ME, Moran K, Feinstein BA, Forscher E, Mustanski B. **Pre-exposure prophylaxis (PrEP) use and condomless anal sex.** *J Acquir Immune Defic Syndr* 2018; **77**:358–364.
11. Oldenburg CE, Nunn AS, Montgomery M, Almonte A, Mena L, Patel RR, *et al*. **Behavioral changes following uptake of HIV preexposure prophylaxis among men who have sex with men in a clinical setting.** *AIDS Behav* 2018; **22**:1075–1079.
12. Volk JE, Marcus JL, Phengrasamy T, Blechinger D, Nguyen DP, Follansbee S, *et al*. **No new HIV infections with increasing use of HIV preexposure prophylaxis in a clinical practice setting.** *Clin Infect Dis* 2015; **61**:1601–1603.
13. Traeger MW, Schroeder SE, Wright EJ, Hellard ME, Cornelisse VJ, Doyle JS, *et al*. **Effects of preexposure prophylaxis for the prevention of HIV infection on sexual risk behavior in men who have sex with men: a systematic review and meta-analysis.** *Clin Infect Dis* 2018; **67**:676–686.
14. Beymer MR, Devost MA, Weiss RE, Dierst-Davies R, Shover CL, Landovitz RJ, *et al*. **Does HIV preexposure prophylaxis use lead to a higher incidence of sexually transmitted infections? A case-crossover study of men who have sex with men in Los Angeles, California.** *Sex Transm Infect* 2018; **94**:457–462.
15. Kojima N, Davey DJ, Klausner JD. **Preexposure prophylaxis for HIV infection and new sexually transmitted infections among men who have sex with men.** *AIDS* 2016; **30**:2251–2252.
16. Liu AY, Cohen SE, Vittinghoff E, Anderson PL, Doblecki-Lewis S, Bacon O, *et al*. **Preexposure prophylaxis for HIV infection integrated with municipal- and community-based sexual health services.** *JAMA Intern Med* 2016; **176**:75–84.
17. Lal L, Audsley J, Murphy DA, Fairley CK, Stoove M, Roth N, *et al*. **Medication adherence, condom use and sexually transmitted infections in Australian preexposure prophylaxis users.** *AIDS* 2017; **31**:1709–1714.
18. Nguyen V-K, Greenwald ZR, Trottier H, Cadieux M, Goyette A, Beauchemin M, *et al*. **Incidence of sexually transmitted infections before and after pre-exposure prophylaxis for HIV.** *AIDS* 2018; **32**:523–530.
19. Marcus JL, Hurley LB, Hare CB, Nguyen DP, Phengrasamy T, Silverberg MJ, *et al*. **Preexposure prophylaxis for HIV prevention in a large integrated healthcare system: adherence, renal safety, and discontinuation.** *J Acquir Immune Defic Syndr* 2016; **73**:540–546.
20. Public Health - Seattle & King County. Preexposure prophylaxis (PrEP) implementation guidelines 2015 [Internet]. 2015. https://www.kingcounty.gov/depts/health/communicable-diseases/hiv-std/patients/~/media/depts/health/communicable-diseases/documents/hivstd/PrEP-implementation-guidelines.ashx [Accessed 11 October 2017]
21. Public Health Seattle & King County. HIV and sexually transmitted infection (STI) screening recommendations for men who have sex with men (MSM) and transgender and non-binary (TG/NB) persons who have sex with men [Internet]. 2018. https://www.kingcounty.gov/depts/health/communicable-diseases/hiv-std/providers/testing-msm-trans.aspx [Accessed 11 May 2018]
22. Stuart EA. **Matching methods for causal inference: a review and a look forward.** *Stat Sci* 2010; **25**:1–21.
23. Lunt M. **Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching.** *Am J Epidemiol* 2014; **179**:226–235.
24. Sainani KL. **Propensity scores: uses and limitations.** *Phys Med Rehabil* 2012; **4**:693–697.
25. Austin PC. **An introduction to propensity score methods for reducing the effects of confounding in observational studies.** *Multivariate Behav Res* 2011; **46**:399–424.
26. Holmes KK, Sparling PF, Stamm WE, Piot P, Wasserheit JN, Corey L, *et al*., editors. *Sexually transmitted diseases*. 4th ed. New York: McGraw Hill; 2008.
27. Kent CK, Chaw JK, Wong W, Liska S, Gibson S, Hubbard G, *et al*. **Prevalence of rectal, urethral, and pharyngeal chlamydia and gonorrhea detected in 2 clinical settings among men who have sex with men: San Francisco, California.** *Clin Infect Dis* 2005; **41**:67–74.
28. Enamorado T, Fifield B, Imai K. **Using a probabilistic model to assist merging of large-scale administrative records.** *Am Political Sci Rev* 2018; **113**:353–371.
29. Public Health - Seattle & King County. HIV and sexually transmitted infection (STI) screening recommendations [Internet]. https://www.kingcounty.gov/depts/health/communicable-diseases/hiv-std/providers/testing-msm-trans.aspx [Accessed 26 November 2018]
30. Public Health Seattle & King County. King county sexually transmitted diseases epidemiology report. 2015. https://www.kingcounty.gov/depts/health/communicable-diseases/hiv-std/patients/epidemiology/~/media/depts/health/communicable-diseases/documents/hivstd/2015-STD-Epidemiology-Report.ashx [Accessed November 2018]
31. Werner RN, Gaskins M, Nast A, Dressler C. **Incidence of sexually transmitted infections in men who have sex with men and who are at substantial risk of HIV infection: a meta- analysis of data from trials and observational studies of HIV preexposure prophylaxis.** *PLoS One* 2018; **13**:1–24.

Supplementary Figure I. Distribution of Propensity Scores among PrEP Users and Non-Users



Supplementary Table 1. Standardized Differences for Propensity Score Covariates, Comparing PrEP Users and Propensity Score-Matched PrEP Non-Users at the Time of Their Initial PHSKC STD Clinic visit (N=1095)[a]

	PrEP Users (N=365)		PrEP Non-Users (N=730)			
Characteristic	**N**	**%**	**N**	**%**	**P-value**	**Standardized Difference**
Age (Mean, SD)	30.6	8.7	30.1	8.6	0.33	6.3
Race/Ethnicity[c]						
White, non-Hispanic	183	50.1	390	53.4	0.11	6.6
Black/African American, non-Hispanic	25	6.8	43	5.9		3.9
Hispanic	90	24.7	137	18.8		14.3
Asian/Pacific Islander, non-Hispanic	41	11.2	76	10.4		2.6
Mixed Race/Other, non-Hispanic[d]	14	2.7	42	4.7		9.0
Unknown, non-Hispanic	12	3.3	42	5.8		11.9
Visit Reason						
Symptoms	198	27.1	100	27.4	0.92	0.6
STI Screening	288	39.5	147	40.3	0.79	1.7
Positive STI Test	115	15.8	47	12.9	0.21	8.2
Partner has Symptoms or STI Diagnosis	212	29.0	109	29.9	0.78	1.8
Health Department Partner Services	18	2.5	12	3.3	0.43	4.9
Contact to Gonorrhea	62	8.5	37	10.1	0.37	5.7
Contact to Chlamydia	56	7.7	32	8.8	0.53	4.0
Contact to Syphilis	28	3.8	26	7.1	0.02	14.5
Symptoms						
Discharge	60	8.2	35	9.6	0.45	4.8
Painful Urination	60	8.2	38	10.4	0.23	7.5
Penile Discomfort	72	9.9	32	8.8	0.56	3.8
Genital Rash	83	11.4	29	7.9	0.08	11.6
Body Rash	64	8.8	23	6.3	0.15	9.3
Anorectal Discomfort	77	10.5	36	9.9	0.73	2.3
Testicular Discomfort	26	3.6	11	3.0	0.64	3.1
Other	82	11.2	38	10.4	0.68	2.6
Sexual Behavior[e]						
New Male Sex Partner, Prior 2 Months	513	70.3	258	70.7	0.89	0.9
Oral Sex, Receptive	575	78.8	286	78.4	0.88	1.0
Oral Sex, Insertive	584	80.0	284	77.8	0.40	5.4
Anal Sex, Receptive	565	77.4	272	74.5	0.29	6.7
With HIV-Positive Partner	81	11.1	46	12.6	0.46	4.7
Condom Use						
N/A	649	88.9	320	87.7	0.49	3.8
Always	23	3.2	8	2.2		5.9
Usually	23	3.2	12	3.3		0.8
Sometimes	16	2.2	14	3.8		9.6
Never	19	2.6	11	3.0		2.5
With HIV-Negative Partner	502	68.8	234	64.1	0.12	9.9
Condom Use						

N/A	232	31.8	133	36.4		9.8
Always	107	14.7	32	8.8		18.4
Usually	192	26.3	97	26.6	0.017	0.6
Sometimes	123	16.8	75	20.5		9.5
Never	76	10.4	28	7.7		9.6
With Unknown Status Partner	169	23.2	91	24.9	0.51	4.2
Condom Use						
N/A	562	77.0	274	75.1		4.5
Always	46	6.3	13	3.6		12.7
Usually	45	6.2	37	10.1	0.006	14.5
Sometimes	41	5.6	31	8.5		11.2
Never	36	4.9	10	2.7		11.4
Anal Sex, Insertive	550	75.3	261	71.5	0.17	8.7
With HIV-Positive Partner	147	20.1	70	19.2	0.71	2.4
Condom Use						
N/A	585	80.1	295	80.8		1.7
Always	56	7.7	15	4.1		15.2
Usually	31	4.2	23	6.3	0.13	9.2
Sometimes	36	4.9	20	5.5		2.5
Never	22	3.0	12	3.3		1.6
With HIV-Negative Partner	519	71.1	240	65.8	0.07	11.5
Condom Use						
N/A	215	29.5	125	34.2		10.3
Always	107	14.7	42	11.5		9.3
Usually	197	27.0	85	23.3	0.06	8.5
Sometimes	119	16.3	76	20.8		11.6
Never	92	12.6	37	10.1		7.8
With Unknown Status Partner	196	26.8	98	26.8	1.0	0.0
Condom Use						
N/A	534	73.2	268	73.4		0.6
Always	49	6.7	14	3.8		12.9
Usually	61	8.4	37	10.1	0.12	6.1
Sometimes	49	6.7	33	9.0		8.6
Never	37	5.1	13	3.6		7.4
Female Sex Partner, Prior 12 Months	33	4.5	23	6.3	0.21	7.9
Female Sex Partner, Prior 2 Months	17	2.3	9	2.5	0.89	0.9
New Female Sex Partner, Prior 2 Months	13	1.8	6	1.6	0.87	1.1
Vaginal Sex, Female Partner, Prior 2 Months	13	1.8	8	2.2	0.64	2.9
Condom Use						
N/A	717	98.2	357	97.8		2.9
Always	3	0.4	2	0.5		2.0
Usually	2	0.3	1	0.3	0.98	0.0
Sometimes	4	0.5	3	0.8		3.3
Never	4	0.5	2	0.5		0.0
Oral Sex, Female Partner, Prior 2 Months						
Gave	14	1.9	8	2.2	0.76	1.9
Received	12	1.6	8	2.2	0.52	4.0
STI Diagnoses[f]						
Pharyngeal Gonorrhea	146	20.0	71	19.5	0.83	1.4

Urethral Gonorrhea	66	9.0	38	10.4	0.47	4.6
Rectal Gonorrhea	221	30.3	108	2.6	0.82	1.5
Pharyngeal Chlamydia	30	4.1	18	4.9	0.53	4.0
Urethral Chlamydia	48	6.6	29	7.9	0.4	5.3
Rectal Chlamydia	152	20.8	87	23.8	0.26	7.2
Early Syphilis	106	14.5	58	15.9	0.55	3.8
Presumptive NGU Diagnosis, Current Visit	33	4.5	17	4.7	0.92	0.7
HIV Risk Behaviors[g]						
Sex with Injecting Drug User	51	7	24	6.6	0.8	1.6
Sex with HIV-POSITIVE Partner	176	24.1	86	23.6	0.84	1.3
Sex with Transgender Person	20	2.7	5	1.4	0.15	9.7
Sex with Anonymous Partner	281	38.5	130	35.6	0.35	6
Sex with Partner from the Internet	441	60.4	224	61.4	0.76	2
Sex at Bath House	156	21.4	74	20.3	0.67	2.7
Injection Drug Use	23	3.2	12	3.3	0.9	0.8
Crack Cocaine Use	9	1.2	6	1.6	0.58	3.4
Erectile Dysfunction Drug Use	93	12.7	44	12.1	0.75	2.1
Popper Use	313	42.9	175	47.9	0.11	10.2
Methamphetamine Use	81	11.1	36	9.9	0.53	4
Exchange Sex[h]	21	2.9	11	3	0.9	0.8

PHSKC, Public Health – Seattle & King County; SD, standard deviation; NGU, non-gonococcal urethritis
[a]Bold denotes variables where any standardized differences are greater than 10%
[b]Standardized differences are reported as the absolute value of the standardized difference comparing the two groups
[c]Categories for race/ethnicity are mutually exclusive
[d]Other includes self-reported other race and Native American and Alaskan Native
[e]Self reported sexual behavior in prior 12 months, unless otherwise specified
[f]In prior 12 months, unless otherwise specified; includes diagnosis at current visit
[g]Selfreported, in prior 12 months
[h]Giving or receiving money or drugs in exchange for sex.