Release bias in accessing medical records in clinical trials: a STAR*D report

T. MICHAEL KASHNER,^{1,2} MADHUKAR H. TRIVEDI,¹ ANNIE WICKER,¹ MAURIZIO FAVA,³
KATHY SHORES-WILSON,¹ STEPHEN R. WISNIEWSKI⁴ & A. JOHN RUSH⁵

1 Department of Psychiatry, The University of Texas Southwestern Medical Center at Dallas, TX, USA

2 Department of Veterans Affairs, Veterans Health Administration Office of Academic Affiliations, Washington, DC, USA

3 Harvard Medical School and Massachusetts General Hospital, Boston, MA, USA

4 Epidemiology Data Center, University of Pittsburgh, Pittsburgh, PA, USA

5 Department of Clinical Sciences, The University of Texas Southwestern Medical Center at Dallas, TX, USA

Key words

major depression, sample selection bias, informed consent

Correspondence

T. Michael Kashner, Department of Psychiatry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9086, USA.
Telephone (+1) 214 648-4608
Fax (+1) 214 648-4612
Email: michael.kashner@va.gov

Received 11 July 2007;
revised 16 June 2008;
accepted 20 November 2008

Abstract

Clinical trials often require subjects to sign medical record releases to allow investigators to measure treatment fidelity, off-protocol care use, and care costs. Little is known, however, if limiting samples to those willing to sign releases impacts external validity. Data came from outpatients with non-psychotic major depressive disorder who enrolled in the multisite Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study. Differences between those who signed ($n = 3116$) and who did not sign ($n = 925$) releases were assessed using logistic regression and two-part, three-level log-transformed regression models, corrected for site clustering and repeated measures. Patients who released records tended to believe care was helpful, were younger, and married. However, release status had little material or consistent associations with patient health outcomes or use of care. With appropriate adjustments to data, requiring patient medical records may pose only minimal challenges to external validity in cost-outcome studies. Copyright © 2009 John Wiley & Sons, Ltd.

Introduction

Investigators conducting clinical trials rely on medical records to assess treatment fidelity, determine use of off-protocol care, and assess costs of care. However, these records can be obtained only if study participants sign

informed consent to allow investigators access. Obtaining these releases may become more difficult as concerns over health information security grow. It thus becomes important to determine how a final analytic sample in a clinical trial may differ whenever subjects who refuse to sign a data release form are excluded from the study sample.

This study assesses the potential selection bias that may occur when access to medical records is a precondition for study enrollment. We focus on patients with major depression as one of the major causes of disability and high health care costs (Rice and Miller, 1993).

Data come from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, described elsewhere (Fava *et al.*, 2003; Rush *et al.*, 2004; Rush *et al.*, 2006b; Trivedi *et al.*, 2006a; Trivedi *et al.*, 2006b). STAR*D was designed to determine which of several treatments were most effective for outpatients with non-psychotic major depressive disorder. The study comprised three- or four-level, sequenced, prospective, randomized clinical trials that involved over 4000 patients at 41 clinics and 400 clinicians.

There are design and sample size advantages to these data. First, STAR*D used a two-step consent procedure in which subjects were first consented to participate in the treatment trial followed by a separate consent to release their medical records. Second, subjects were adult outpatients ranging in age from 18 to 75 years and were selected from 41 clinical sites throughout the United States in psychiatric and primary care settings and in public and private sectors. All patients satisfied DSM-IV (American Psychiatric Association, 1994) criteria for non-psychotic major depressive disorder, had depressive symptoms of 14 or higher on the 17-item Hamilton Rating Scale for Depression (HRSD₁₇) (Hamilton, 1960, 1967), and did not have a primary diagnosis of obsessive compulsive or an eating disorder at baseline. Third, to maintain a large, representative sample, STAR*D developed 'equipoise' stratified randomized design (Lavori *et al.*, 2001) that allowed subjects to refuse randomization to treatments they found unacceptable. This permitted investigators to combine the broad inclusion/exclusion criteria of an effectiveness study with the randomized design of an efficacy trial.

Consent agreements

Investigators can only access medical records if study participants signed an informed consent agreement approved by an independent Institutional Review Board. These agreements specify study purposes, procedures, risks, and potential benefits, explain why access to records is necessary, and describe the scope of services and encounter dates, how health care providers or third-party plans are identified and contacted, and the procedures and timelines for data collection and transfer, storage, de-identification, and final disposition of data when study ends.

The administrative process to obtain medical records is further complicated because the licensed practitioners, hospitals, and health plans who hold the medical records are often not parties to the original consent agreement. Thus, release forms must often include specific language that had been supplied by the provider. To protect patient privacy, STAR*D further required all references to patients participating in a mental health study be removed before a signed release form could be submitted to a provider.

Methods

Enrollment

To enter the protocol trial, all STAR*D participants signed informed consent during face-to-face interviews with a clinical research coordinator. After signing informed consent, participants were then asked to sign a form that would release medical records to STAR*D investigators for the expressed purpose of determining patient use and cost of health care. Patients were told their decision to sign a release would not impact their participation in the clinical trial. STAR*D participants who agreed to release their medical records were called 'releasers,' while those declining were called 'non-releasers.'

All patients entered level 1 treatment (citalopram) and remained until achieving remission up to a maximum of 12–14 weeks. Those who did not achieve remission received additional treatment for up to a total three to four treatment levels. Patients who achieved remission entered a one-year follow-up phase. Those who never achieved remission at the end of the final treatment level were exited from the study. Clinical remission was based on the judgment of the treating physician using the 30-item, clinician-rated Inventory of Depressive Symptomatology (IDS-C₃₀) (Rush *et al.*, 1996; Rush *et al.*, 2000; Trivedi *et al.*, 2004).

Data collection

As described elsewhere (Fava *et al.*, 2003; Rush *et al.*, 2004; Rush *et al.*, 2006b; Trivedi *et al.*, 2006a; Trivedi *et al.*, 2006b), demographic, education, earnings, employment, and health insurance information were collected at baseline from face-to-face and telephone interviews with study patients. Noting the importance of patient health beliefs on treatment efficacy (Cooper *et al.*, 2003), patients also described their care attitudes on a five-point scale: 'If I can get the help I need from a doctor, I believe that I will be much better able: (a) to make important decisions that affect my life and those of my family? and (b) to enjoy

things that interest me?' These questions explained 75% of the total variance from an original 10-item questionnaire (Kashner *et al.*, 2003; Kashner *et al.*, 2006; Rush *et al.*, 2003a). For the present analyses, responses were dichotomized reflecting agree versus neutral or disagree. Patients were also asked to rate: '...the current overall impact of your family and friends on your condition...' on a seven-point scale, ranging from very helpful, moderately helpful, minimally helpful, neutral, and minimally more difficult, moderately more difficult, and much more difficult. Responses were dichotomized to reflect helpful versus neutral or not helpful.

Level 1 treatment outcomes were based on HRSD₁₇ (Hamilton, 1960, 1967) that was administered by telephone at the end of level 1, and on the written, patient self-reported 16-item Quick Inventory of Depressive Symptomatology (QIDS-SR₁₆) (Rush *et al.*, 2000; Rush *et al.*, 2003b; Rush *et al.*, 2006a; Trivedi *et al.*, 2004), administered at baseline and at each STAR*D clinic visit. Outcomes were computed as: (1) remission defined by a HRSD₁₇ score of seven or less at the end of level 1, (2) remission defined by a QIDS-SR₁₆ score of five or less, and (3) response defined by a reduction in QIDS-SR₁₆ score from baseline of 50% or more.

Information was also collected using an Interactive Voice Response (IVR) system where patients called into a centralized computer and answer questions using their telephone keypad. Described elsewhere (Kashner *et al.*, 2009b), collected information included baseline SF-12 mental and physical functioning (Ware *et al.*, 1996), family size, health insurance status, and US Medicaid eligibility, and patient use of health care measured in 90-day intervals. Use of care included the number of outpatient clinic visits, emergency room visits, or inpatient days stayed (treatment setting) based on whether the care was depression-related, other-psychiatric-related, or for general medical problems (clinical purpose). Scripts were adapted from the Utilization and Cost Inventory (Kashner *et al.*, 2009a) and the Utilization and Cost Questionnaire (Kashner *et al.*, 1999a; Kashner *et al.*, 1999b; Kashner *et al.*, 2006; Rush *et al.*, 2003a).

For this study, use of care was summarized in 90-day intervals by treatment setting and clinical purpose, as described elsewhere (Kashner *et al.*, in press). Briefly, nine services were defined for each of the three treatment settings and three clinical purposes. For each service, care use was presented as a dichotomous use variable that assumed the value of one if the patient used service, and zero otherwise. For patients who used service, the volume of care was computed in 'dollars' by multiplying service counts (visit, days) by a unit cost per count. Unit cost

weights were estimated from STAR*D provider records. These records contained patient diagnoses and information about both outpatient procedures in terms of Current Procedural Terminology (CPT) (American Medical Association, 2004) and inpatient days by diagnosis-related groups (DRG) (St Anthony, 2004). This information was used to classify care into discrete service counts (visits, days) classified into one of the nine service categories. The cost for each count was computed by multiplying the number of included CPT procedures or DRG days by a respective market based rate. These rates were taken from the Department of Veterans Affairs Reasonable Charges program (version 2.4, 70 (68) F.R. 18465 April 11, 2005) (Kashner *et al.*, 2006; Kashner *et al.*, 2009a; Kashner *et al.*, 2009b; Kashner *et al.*, in press) representing the 80th percentile of paid transaction rates in the United States. Unit costs were computed from a mean cost per count, by service, that was adjusted to reflect the STAR*D sample mean. Adjustments were computed from regressions with the patient-sample mean-centered covariates that included African-American and Hispanic status, gender, physical and mental functioning, Medicaid and private insurance, student and employment status, age, number in household, marital status, high school graduate, baseline QIDS-SR₁₆, Axis I and Axis III diagnoses, and whether the patient had a family history of depression. Variation in costs among patients was intended to reflect the mix and volume of care received, rather than local market prices and production efficiencies.

Analyses

Simple comparisons were based on Pearson chi-square with 2×2 containing continuity corrections for discrete cases, and two-tailed *t*-test with unequal variances assumed for continuous cases. Predictors of signed releases and the impact of signed releases on dichotomous health outcomes were computed using two-level logistic regressions corrected to account for facility nesting (random effect), and adjusted to reflect mean-centered core covariates. Based on a conceptual framework by Anderson *et al.* (1968), core covariates included patient care needs (QIDS-SR₁₆ score assessed at baseline, patient age), predisposing factors (graduated high school, Hispanic, African-American, gender), and enabling characteristics (married, enrolled in private health insurance plan, and employed). When assessing health outcomes and use of care, core covariates were expanded to include other need-related variables (SF-12 mental and SF-12 physical scales) and call order (first call = 1, second call = 2, ...). Values for costs were log-transformed to account

for skewed distributions. The association of release status on costs were assessed as two-part, three-level, mixed models (Bryck and Raudenbush, 1992) corrected for facility nesting and repeated calls within patient. The impact of release status on costs were assessed for the mean STAR*D subject as a cost ratio (adjusted cost for releasers divided by adjusted cost for non-releasers) calculated by exponentiating the fixed, level-2 coefficient for the release status indicator variable in the log-cost equations, and adjusted for heteroskedasticity (Manning, 1998). Thus, a cost ratio of one indicates no differences in costs between releasers and non-releasers. Finally, significance tests were based on robust estimates of standard errors to account for more complex error distributions.

Results

Table 1 displays baseline characteristics, the volume of IVR calls, and health care use for the 4041 patients participating in the STAR*D clinical trial. Values are broken down for those who signed ($n = 3116$, 77.1%) and who did not sign ($n = 925$, 22.9%) a medical release form.

Table 2 describes the association between baseline characteristics and release status after adjusting for core covariates and facility nesting. Patients were more likely to release medical records if they were younger, were married or living with a spouse, and agreed with the statement that care for their depression would help them with decision-making and with enjoying things of interest. However, patient attitudes about whether they perceive their families as helpful with their condition, baseline health, and demographic factors including ethnicity and race, gender, and socio-economic status, were not materially or statistically significantly associated with whether patients signed releases.

Table 3 describes how release status was associated with health outcomes and with use of care after adjusting for covariates and facility nesting. (The 95% confidence intervals are shown in square brackets in text.) Overall, release status had small and statistically insignificant impacts on either outcomes of care (treatment response and remission) or on patient reported use of care.

Thesedata did not show significant associations between willingness to sign medical releases and patient response to depression treatment. However, patients who signed releases incurred -15% [-26%, -3%] fewer costs in other-mental health services for non-depression purposes.

Finally, to test the construct validity of our analytic framework, we examined the association between patient characteristics and clinical need covariates on patient use of psychiatric and general medical care. African-

Americans reported -30% ([-45%, -10%], $t = 2.81$, $p = 0.005$) and Hispanics reported -47% ([-63%, -25%], $t = 3.54$, $p = 0.001$) fewer services than their respective counterparts. Older patients reported using more services by 14% ([5%, 23%], $t = 3.09$, $p = 0.002$) for every 10 years of age. Patients also used more care by 4% ([1%, 7%], $t = 2.81$, $p = 0.005$) per unit baseline QIDS-SR₁₆ score, by 96% ([75%, 122%], $t = 11.75$, $p < 0.001$) per 10 unit drop of SF-12 physical functioning, by 35% ([14%, 61%], $t = 3.45$, $p = 0.001$) per 10 unit drop of SF-12 mental functioning, by 129% ([65%, 219%], $t = 4.93$, $p < 0.001$) for a high school graduate or more, and by 72% ([34%, 121%], $t = 4.30$, $p < 0.001$) for subjects with private health insurance. Patients with more depressive symptoms were more likely to seek non-depression-related mental health care (by 4% per unit QIDS-SR₁₆, [3%, 6%], $t = 5.21$, $p < 0.001$), while patients in poorer physical functioning sought more general medical care (by 23% per 10 unit drop in SF-12 physical, [19%, 27%], $t = 14.53$, $p < 0.001$).

Discussion

STAR*D provided a large sample size and a unique design to test the impact on sample generalizability when study participants are required to sign release forms. Our findings suggest that requiring patients to sign releases before entering a trial would not substantively bias the final sample to favor either positive or negative treatment outcomes, or higher or lower treatment costs. However, it would favor patient samples with favorable attitudes towards care, younger ages, and supporting families.

Prior studies have shown adverse reactions to outcome surveys for patients who received mental health treatment (Zimmerman *et al.*, 1994), with severe psychiatric symptom scores (Henderson and Jorm, 1990), and with mental illness symptomatology (Boothroyd, 2000). This study suggests that patient reluctance to offer access was more related to spousal support, age, and attitudes towards the effectiveness of care than baseline health status or depression symptom. This interpretation runs counter to studies on dementia that showed greater adverse response to patient surveys when respondents had poorer cognitive assessments (Jorm *et al.*, 1994). Differences in findings may reflect a STAR*D sample that included only patients willing to participate in a clinical trial.

There are limitations to the study. The study sample included only those patients who agreed to the trial. Thus, results may not necessarily reflect a broader population of patients with major depression. Covariates were adjusted only for linear effects. Selection bias will

Table 1 Patient characteristics, by medical release status

	Non releasers (<i>n</i> = 925) ¹		Releasers (<i>n</i> = 3116) ¹			
	<i>n</i>	Percentage total		Percentage total		
	Mean	Standard deviation	<i>n</i>	Standard deviation	Test-statistic ²	<i>p</i> -value
<i>Demographic</i>						
Ethnicity (<i>n</i> = 4036)						
White	739	80.0%	2316	74.4%	$\chi^2(5) = 16.7$	0.005
African American	124	13.4%	585	18.8%		
Asian American	20	2.2%	50	1.6%		
Native American	8	0.9%	32	1.0%		
Pacific Islander	5	0.5%	18	0.6%		
Multi-racial	28	3.0%	111	3.6%	$\chi^2(1) = 32.7$	<0.001
Hispanic ³	65	7.0%	442	14.2%		
Female (<i>n</i> = 4041)	598	64.6%	1934	62.1%	$\chi^2(1) = 1.9$	0.17
Age (<i>n</i> = 4041)	40.4 year	±18.1	41.4 year	±15.0	$t(1323) = 1.6$	0.12
Age (<i>n</i> = 4041)						
<25	139	15.0%	379	12.2%	$\chi^2(6) = 16.5$	0.011
25–4	220	23.8%	782	25.1%		
35–44	240	25.9%	686	22.0%		
45–54	189	20.4%	738	23.7%		
55–64	101	10.9%	394	12.6%		
65–74	32	3.5%	129	4.1%		
75+	4	0.4%	8	0.3%		
<i>Household/residence</i>						
Residence type (<i>n</i> = 4035)						
Detached	472	51.1%	1637	52.6%	$\chi^2(4) = 12.5$	0.014
Row/town	41	4.4%	134	4.3%		
Mobile	33	3.6%	188	6.0%		
Apartment	353	38.2%	1055	33.9%		
Other	25	2.7%	97	3.1%		
Years at current residence (<i>n</i> = 3351)	5.6	±8.2	5.4	±8.0	$t(1311) = 0.6$	0.52
Marital status (<i>n</i> = 4035)						
Married	282	30.5%	1050	33.8%	$\chi^2(3) = 4.1$	0.25
Separated/divorced	237	25.6%	800	25.7%		
Widowed	31	3.4%	97	3.1%		
Other	374	40.5%	1164	37.4%		
Household size (<i>n</i> = 4032)	2.8	±3.3	2.9	±1.7	$t(1074) = 0.6$	0.56
<i>Socio-economic status</i>						
Income per month						
Household income (<i>n</i> = 3819)	\$2369	±3575	\$2178	±2955	$t(1270) = 1.5$	0.15

Table 1 Continued

	Non releasers (<i>n</i> = 925) ¹		Releasers (<i>n</i> = 3116) ¹		Test-statistic ²	<i>p</i> -value
	<i>n</i> Mean	Percentage total Standard deviation	<i>n</i> Mean	Percentage total Standard deviation		
Earned income (<i>n</i> = 3751)	\$1480	±2788	\$1401	±2488	<i>t</i> (1335) = 0.8	0.45
Number with earnings from employment (<i>n</i> = 3751)	504	60.8%	1749	59.8%	$\chi^2(1) = 0.2$	0.64
Employment (<i>n</i> = 4034)						
Retired	43	4.7%	191	6.1%		
Unemployed						
Not looking	155	16.8%	686	22.1%		
Looking	151	16.3%	497	16.0%	$\chi^2(5) = 17.9$	0.003
Employed						
Part-time	131	14.2%	380	12.0%		
Full-time	398	43.1%	1204	38.7%		
Self	46	5.0%	152	4.9%		
Education (<i>n</i> = 4034)						
None	77	8.3%	439	14.1%		
High school graduate	372	40.3%	1291	41.5%		
GED	64	6.9%	242	7.8%		
Associate degree	124	13.4%	406	13.1%	$\chi^2(6) = 35.7$	<0.001
College degree	191	20.7%	495	15.9%		
Master degree	71	7.7%	175	5.6%		
Doctorate/professor	25	2.7%	62	2.0%		
Student status (<i>n</i> = 4035)						
None	757	81.9%	2685	86.3%		
Part-time	69	7.5%	193	6.2%	$\chi^2(2) = 11.8$	0.003
Full-time	98	10.6%	233	7.5%		
Volunteer status (<i>n</i> = 4031)						
None	801	86.9%	2650	85.3%		
Part-time	119	12.9%	434	14.0%	$\chi^2(2) = 4.2$	0.12
Full-time	2	0.2%	24	0.8%		
Insurance status						
Medicare enrollee (<i>n</i> = 3949)	52	5.8%	262	8.6%	$\chi^2(1) = 7.3$	0.007
Medicaid enrollee (<i>n</i> = 3929)	91	10.1%	405	13.4%	$\chi^2(1) = 6.7$	0.010
Private insurance enrollee (<i>n</i> = 3954)	510	56.4%	1512	49.6%	$\chi^2(1) = 12.5$	<0.001
Perception of health care						
Agree impact of family/ friends is helpful (<i>n</i> = 4021)	518	56.4%	1827	58.9%	$\chi^2(1) = 1.7$	0.20

Table 1 *Continued*

	Non releasers (<i>n</i> = 925) ¹		Releasers (<i>n</i> = 3116) ¹			
	<i>n</i>	Percentage total Standard deviation	<i>n</i>	Percentage total Standard deviation	Test-statistic ²	<i>p</i> -value
	Mean		Mean			
Agree care will improve decision making functioning (<i>n</i> = 4023)	823	89.7%	2827	91.1%	$\chi^2(1) = 1.5$	0.22
Agree care will make things more enjoyable (<i>n</i> = 4023)	830	90.4%	2860	92.1%	$\chi^2(1) = 2.5$	0.12
<i>Health status</i>						
Family history of depression ⁴ (<i>n</i> = 4000)	515	56.6%	1665	53.9%	$\chi^2(1) = 2.0$	0.16
QIDS-SR ⁵ (<i>n</i> = 3790)	15.2	±5.3	15.4	±5.2	$t(1344) = 0.8$	0.40
Mental status ⁶ (<i>n</i> = 3790)	26.0	±9.1	27.2	±8.9	$t(1330) = 3.3$	0.001
Physical status ⁶ (<i>n</i> = 3790)	51.5	±11.5	48.7	±12.0	$t(1409) = 6.2$	<0.001
Medical/psychiatric leave (<i>n</i> = 4031)	64	6.9%	248	8.0%	$\chi^2(1) = 1.0$	0.32
<i>IVR calling status</i>						
Initializing ⁷ first call (<i>n</i> = 4041)	844	91.2%	2945	94.5%	$\chi^2(1) = 12.5$	<0.001
Mean number calls (<i>n</i> = 3789)	2.4	±1.6	2.7	±1.7	$t(1504) = 0.9$	0.35
Number calls among initializers ⁷ (<i>n</i> = 3789)						
1 call	371	44.0%	1097	37.2%	$\chi^2(7) = 18.5$	0.010
2 calls	164	19.4%	574	19.5%		
3 calls	104	12.3%	361	12.3%		
4 calls	66	7.8%	292	9.9%		
5 calls	92	10.9%	409	13.9%		
6 calls	43	5.1%	183	6.2%		
7 calls	4	0.5%	28	1.0%		
8 calls	0	0.0%	1	0.0%		
<i>Use of care per IVR call</i> (<i>n</i> = 9864) ⁸						
Outpatient care						
Depress-related						
Number reporting	957	47.4%	3772	48.1%	$\chi^2(1) = 0.33$	0.57
Volume – visits	4.25	±5.71	4.02	±4.48	$t(1271) = 1.1$	0.26
Other-psychiatric						
Number reporting	190	9.4%	858	10.9%	$\chi^2(1) = 3.8$	0.050
Volume – visits	3.52	±6.47	3.09	±4.99	$t(241) = 0.9$	0.38
General-medical						
Number reporting	632	31.3%	2648	33.8%	$\chi^2(1) = 4.4$	0.036
Volume – visits	2.83	±2.88	3.08	±3.99	$t(1277) = 1.8$	0.071

Table 1 *Continued*

	Non releasers (<i>n</i> = 925) ¹		Releasers (<i>n</i> = 3116) ¹		Test-statistic ²	<i>p</i> -value
	<i>n</i> Mean	Percentage total Standard deviation	<i>n</i> Mean	Percentage total Standard deviation		
Emergency room						
Depress-related						
Number reporting	95	4.7%	351	4.5%	$\chi^2(1) = 0.1$	0.71
Volume – visits	1.25	±0.89	1.48	±2.06	$t(362) = 1.6$	0.11
Other-psychiatric						
Number reporting	26	1.3%	106	1.4%	$\chi^2(1) = 0.0$	0.91
Volume – visits	1.69	±1.01	1.45	±0.85	$t(34) = 1.1$	0.27
General-medical						
Number reporting	174	8.6%	763	9.7%	$\chi^2(1) = 2.2$	0.14
Volume – visits	1.55	±1.10	1.69	±1.56	$t(350) = 1.5$	0.14
Hospital stay						
Depress-related						
Number reporting	35	1.7%	131	1.7%	$\chi^2(1) = 0.0$	0.93
Volume – days	5.00	±3.04	5.27	±5.43	$t(98) = 0.4$	0.70
Other-psychiatric						
Number reporting	11	0.5%	31	0.4%	$\chi^2(1) = 0.5$	0.47
Volume – days	3.64	±2.42	7.32	±8.44	$t(39) = 2.2$	0.035
General-medical						
Number reporting	50	2.5%	261	3.3%	$\chi^2(1) = 3.6$	0.059
Volume – days	5.34	±7.26	5.69	±9.41	$t(84) = 0.3$	0.77

¹Number in parentheses is number of patients reporting.

²Discrete cases based on Pearson chi-square (χ^2), with 2×2 containing continuity correction. Continuous cases based on two-tailed *t*-test with unequal variance assumed.

³Includes both Black and White races.

⁴Family history of major depression including parent, sibling, or child.

⁵Quick Inventory of Depressive Symptomatology – patient self-report.

⁶Based on the Medical Outcome Study SF-12.

⁷Initializers are study subjects who initiated and completed at least one IVR call.

⁸Per call, with volume reported in visits or days given respondent indicated use.

Table 2 Adjusted¹ odds ratios of selected predictors that a STAR*D subject will sign a medical release form

	exp β	95% Confidence interval		Wald <i>t</i>	<i>p</i> -value
Attitudes					
Care believed helpful for decision	1.33	1.03	1.72	2.17	0.030
Care believed helpful enjoy things	1.35	1.06	1.72	2.46	0.014
Family believed helpful	1.08	0.97	1.21	1.39	0.164
Baseline health					
Baseline QIDS-SR ₁₆	1.01	0.99	1.02	0.75	0.451
Family history depression	0.99	0.87	1.14	0.12	0.907
Physical functioning per 10 pts	0.95	0.90	1.00	1.81	0.070
Mental functioning per 10 pts	1.02	0.94	1.10	0.37	0.713
Demographic characteristics					
Age per 10 years	0.95	0.90	1.00	2.06	0.039
Male	1.00	0.84	1.19	0.00	0.999
Hispanic	1.08	0.85	1.38	0.64	0.523
African-American	1.01	0.84	1.22	0.12	0.905
Married	1.21	1.03	1.41	2.31	0.021
Spouse living with patient	1.28	1.06	1.55	2.51	0.012
Size of household	0.98	0.94	1.02	0.94	0.349
Student	0.78	0.61	1.00	1.92	0.054
Volunteer work	1.22	0.94	1.58	1.49	0.135
High School graduate, or higher	0.90	0.67	1.20	0.75	0.454
With private insurance	1.00	0.86	1.16	0.01	0.995
With Medicaid	1.04	0.51	2.13	0.11	0.915
Employed full or part time	1.01	0.87	1.18	0.17	0.866
Household income per 3000	1.01	0.93	1.09	0.28	0.778
Employment income per 3000	1.03	0.92	1.14	0.49	0.622

¹ Estimates of predictors that were non-core covariates were adjusted for core covariates. Estimates for predictors that were core covariates were adjusted to reflect the other core covariates. Core covariates included: age (years) Hispanic and African-American status, gender, marital status (married), education (high school graduate or higher), private insurance status, employment status, and baseline QIDS-SR₁₆.

Table 3 Adjusted¹ estimates of effect of release status on patient health outcomes and on use of care per call for STAR*D subjects²

	Exp β	95% Confidence interval		Wald t	p -value
Treatment outcome					
Response to treatment ³	0.99	0.94	1.03	0.53	0.598
Remission based on HDRS ₁₇ ⁴	1.16	0.93	1.45	1.31	0.190
Remission based on QIDS-SR ₁₆ ⁵	1.05	0.84	1.33	0.43	0.666
Utilization of care					
All care					
Total use ⁶	1.18	0.89	1.55	1.16	0.248
Proportion patients used ⁷	1.10	0.96	1.25	1.40	0.161
Cost per user ⁸	1.03	0.93	1.14	0.54	0.587
Depression-related mental health					
Proportion patients used	1.08	0.95	1.22	1.14	0.253
Cost per user	1.02	0.93	1.12	0.49	0.627
Other mental health					
Proportion patients used	0.96	0.81	1.12	0.55	0.584
Cost per user	0.85	0.74	0.97	2.35	0.019
General medical care					
Proportion patients used	1.13	0.99	1.29	1.86	0.062
Cost per user	1.03	0.97	1.09	0.91	0.366

¹Adjusted for: IVR call order (=1 for first call, =2 for second call), patient age, Hispanic status, African-American status, gender (male), marital status (married versus not married), education (high school or greater versus less than high school), private insurance status (yes/no), employment status (full or part-time versus not employed), baseline QIDS-SR₁₆, baseline SF-12 mental, and SF-12 physical functioning.

²For *n* = 3789 patients making at least one of 9864 IVR calls.

³Response is defined as a reduction in QIDS-SR₁₆ score of 50% or more from baseline.

⁴Remission based on HDRS₁₇ exit score of seven or less.

⁵Remission based on QIDS-SR₁₆ score at endpoint of five or less.

⁶Measured as the ratio of the total adjusted costs of care for releasers divided by the total adjusted cost of care for non-releasers of medical records.

⁷Odds ratio, or the adjusted odds that a releaser uses care divided by the adjusted odds that a non-releasers uses care.

⁸Ratio of costs, or the adjusted cost of releasers who uses care divided by the adjusted cost of non-releasers who uses care. Here, ratio = 1 reflects no differences, ratio > 1 indicates releasers cost more, and ratio < 1 indicates releasers cost less than non-releasers. Costs are computed for any given 90-day period from service information provided by respondents during IVR calls.

ultimately depend on the skills of individual investigators to enroll study patients. How presentation style may influence willingness to sign a release was not explored.

While these results are essentially negative, the large sample size and equipoise design of the STAR*D trial, the small differences between releasing and non-releasing patients compared to traditional predictors, and the good model fit of traditional predictors with these cost and

outcomes data, suggest that requiring medical releases for patients with depression who are otherwise willing to enter a clinical trial, may not introduce materially large selection biases when assessing health outcomes or care costs. At least with these data, with appropriate adjustments researchers may require access to medical records in a clinical trial while creating only minimal challenges to external validity in cost-outcome studies.

Acknowledgments

This project has been funded by funds from the National Institute of Mental Health, National Institutes of Health, under Contract N01-MH-90003, A.J. Rush MD, P.I. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services or the Department of Veterans Affairs, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Declaration of interests statement

This project has been funded by funds from the National Institute of Mental Health, National Institutes of Health.

References

- American Medical Association. (2004) *Current Procedural Terminology*, St Anthony.
- American Psychiatric Association. (1994) *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)*, American Psychiatric Association Press.
- Andersen R., Odin W., Anderson B.S. (1968) Perception of and response to symptoms of illness in Sweden and the United States. *Medical Care*, **6**(1), 18–30.
- Boothroyd R.A. (2000) The impact of research participation on adults with severe mental illness. *Mental Health Service Research*, **2**(4), 213–221.
- Bryck A.S., Raudenbush S.W. (1992) *Hierarchical Linear Models*, Sage Publications.
- Cooper L.A., Gonzales J.J., Gallo J.J., Rost K.M., Meredith L.S., Rubenstein L.V., Wang N.Y., Ford D.E. (2003) The acceptability of treatment for depression among African-American, Hispanic, and white primary care patients. *Medical Care*, **41**(4), 479–489.
- Fava M., Rush A.J., Trivedi M.H., Nierenberg A.A., Thase M.E., Sackeim H.A., Quitkin F.M., Wisniewski S., Lavori P.W., Rosenbaum J.F., Kupfer D.J. (2003) Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatric Clinics of North America*, **26**(2), 457–494, PII S0193-953X(02)00107-7
- Hamilton M. (1960) A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry*, **23**, 56–62.
- Hamilton M. (1967) Development of a rating scale for primary depressive illness. *British Journal of Social & Clinical Psychology*, **6**, 278–296.
- Henderson A.S., Jorm A.F. (1990) Do mental health surveys disturb? *Psychological Medicine*, **20**, 721–724.
- Jorm A.F., Henderson A.S., Scott R., Mackinnon A.J., Korten A.E., Christensen H. (1994) Do mental health surveys disturb? Further evidence. *Psychological Medicine*, **24**, 233–237.
- Kashner T.M., Carmody T.J., Suppes T., Rush A.J., Crismon M.L., Miller A.L., Toprac M., Trivedi M. (2003) Catching up on health outcomes: the Texas Medication Algorithm Project. *Health Service Research*, **38**(1 part 1), 311–331.
- Kashner T.M., Rush A.J., Altshuler K.Z. (1999a) Measuring costs of guideline-driven mental health care: the Texas Medication Algorithm Project. *Journal of Mental Health Policy and Economics*, **2**(3), 111–121.
- Kashner T.M., Rush A.J., Crismon M.L., Toprac M., Carmody T.J., Miller A.L., Trivedi M.H., Wicker A., Suppes T. (2006) An empirical analysis of cost outcomes of the Texas Medication Algorithm Project. *Psychiatric Services*, **57**(5), 648–659.
- Kashner T.M., Stensland M.D., Lind L., Wicker A., Rush A.J., Golden R.M., Henley S.S. (2009a) Measuring use and cost of care for patients with mood disorders: the Utilization and Cost Inventory. *Medical Care*, **47**(2), 184–190.
- Kashner T.M., Suppes T., Rush A.J., Altshuler K.Z. (1999b) Measuring use of outpatient care among mentally ill individuals: a comparison of self-reports and provider records. *Evaluation and Program Planning*, **22**, 31–39.
- Kashner T.M., Trivedi M.H., Wicker A., Fava M., Greist J.H., Mundt J.C., Shores-Wilson K., Rush A.J., Wisniewski S. R. (2009b) Voice response system to measure healthcare costs: a STAR*D report. *American Journal of Managed Care*, **15**(3), 153–162.
- Kashner T.M., Trivedi M.H., Wicker A., Fava M., Wisniewski S.R., Rush A.J. (in press) The impact of non-clinical factors on care use for patients with depression: A STAR*D report. *CNS Neuroscience and Therapeutics*.
- Lavori P.W., Rush A.J., Wisniewski S.R., Alpert J., Fava M., Kupfer D.J., Nierenberg A., Quitkin F.M., Sackeim H.A., Thase M.E., Trivedi M. (2001) Strengthening clinical effectiveness trials: equipoise-stratified randomization. *Biological Psychiatry*, **50**, 792–801, PII S0006-3223(01)01223-9
- Manning W.G. Jr (1998) The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics*, **17**, 283–295.
- Rice D.P., Miller L.S. (1993) The economic burden of affective disorders. *Advances in Health Economics and Health Service Research*, **14**, 37–53.
- Rush A.J., Bernstein I.H., Trivedi M.H., Carmody T.J., Wisniewski S., Mundt J.C., Shores-Wilson K., Biggs M.M., Woo A., Nierenberg A.A., Fava M. (2006a) An evaluation of the Quick Inventory of Depressive Symptomatology and the Hamilton Rating Scale for Depression: a Sequenced Treatment Alternatives to Relieve Depression trial report. *Biological Psychiatry*, **59**, 493–501, Epub 2005 September 30, DOI: 10.1016/j.biopsych.205.08.022
- Rush A.J., Carmody T.J., Reimtz P.E. (2000) The Inventory of Depressive Symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. *International Journal of Methods in Psychiatric Research*, **9**, 45–59.

- Rush A.J., Crismon M.L., Kashner T.M., Toprac M.G., Carmody T.J., Trivedi M.H., Suppes T., Miller A.L., Biggs M.M., Shores-Wilson K., Witte B.P., Shon S.P., Rago W. V., Altshuler K.Z. (2003a) Texas Medication Algorithm Project, phase 3 (TMAP-3): rationale and study design. *Journal of Clinical Psychiatry*, **64**(4), 357–369.
- Rush A.J., Fava M., Wisniewski S.R., Lavori P.W., Trivedi M.H., Sackeim H.A., Thase M.E., Nierenberg A.A., Quitkin F.M., Kashner T.M., Kupfer D.J., Rosenbaum J.F., Alpert J., Stewart J.W., McGrath P.J., Biggs M.M., Shores-Wilson K., Lebowitz B.D., Ritz L., Niederhe G. (2004) Sequenced Treatment Alternatives to Relieve Depression (STAR*D): rationale and design. *Control Clinical Trials*, **25**, 119–142, DOI: 10.1016/S0197-2456(03)00112-0
- Rush A.J., Gullion C.M., Basco M.R., Jarrett R.B., Trivedi M.H. (1996) The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychological Medicine*, **26**, 477–486.
- Rush A.J., Trivedi M.H., Ibrahim H.M., Carmody T.J., Arnow B., Klein D.N., Markowitz J.C., Ninan P.T., Kornstein S., Manber R., Thase M.E., Kocsis J.H., Keller M.B. (2003b) The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, **54**, 573–583, Erratum, p. 585, DOI: 10.1016/S0006-3223(03)01866-8
- Rush A.J., Trivedi M.H., Wisniewski S.R., Stewart J.W., Nierenberg A.A., Thase M.E., Ritz L., Biggs M.M., Warden D., Luther J.F., Shores-Wilson K., Niederhe G., Fava M. (2006b) Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *New England Journal of Medicine*, **354**(12), 1231–1242.
- St Anthony. (2004) *St Anthony's DRG Guidebook*, St Anthony.
- Trivedi M.H., Fava M., Wisniewski S.R., Thase M.E., Quitkin F., Warden D., Ritz L., Nierenberg A.A., Lebowitz B.D., Biggs M.M., Luther J.F., Shores-Wilson K., Rush A.J. (2006a) Medication augmentation after the failure of SSRIs for depression. *New England Journal of Medicine*, **354**(12), 1243–1252.
- Trivedi M.H., Rush A.J., Ibrahim H.M., Carmody T.J., Biggs M.M., Suppes T., Crismon M.L., Shores-Wilson K., Toprac M.G., Dennehy E.B., Witte B., Kashner T.M. (2004) The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychological Medicine*, **34**, 73–82, DOI: 10.1017/S0033291703001107
- Trivedi M.H., Rush A.J., Wisniewski S.R., Nierenberg A.A., Warden D., Ritz L., Norquist G., Howland R.H., Lebowitz B., McGrath P.J., Shores-Wilson K., Biggs M.M., Balasubramani G.K., Fava M. (2006b) Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *American Journal of Psychiatry*, **163**(1), 28–40.
- Ware J.E., Kosinski M., Keller S.D. (1996) A 12-item Short-Form Health Survey. Construction of scales and preliminary tests of reliability and validity. *Medical Care*, **34**(3), 220–233.
- Zimmerman M., Farber N.J., Hartung J., Lush D.T., Kuzma M.A. (1994) Screening for psychiatric disorders in medical patients: a feasibility and patient acceptance study. *Medical Care*, **32**(6), 603–608.