2014-08-18

Comparative Effectiveness of Lithium and Valproate for Suicide Prevention and Associations With Nonsuicide Mortality: A Dissertation

Eric G. Smith
University of Massachusetts Medical School

COMPARATIVE EFFECTIVENESS OF LITHIUMAND VALPROATE FOR SUICIDE PREVENTION AND ASSOCIATIONS WITH NONSUICIDE MORTALITY

A Dissertation Presented

By

Eric Graham Smith, M.D., M.P.H.

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 18, 2014

Millennium Program
Clinical and Population Health Research

COMPARATIVE EFFECTIVENESS OF LITHIUM AND VALPROATE FOR SUICIDE PREVENTION AND ASSOCIATIONS WITH NONSUICIDE MORTALITY

A Dissertation Presented by

Eric Graham Smith, M.D., M.P.H.

The Signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

_____
Jeroan Allison, M.D., M.P.H., Thesis Advisor

_____
Terry Field, D.Sc., Member of Committee

_____
Carl Fulwiler, M.D., Ph.D., Member of Committee

_____
Mary Jo V. Pugh, Ph.D., R.N., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirement of the Dissertation Committee

_____
Becky Briesacher, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduate requirements of the school.

_____
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Millennium Program
Clinical and Population Health Research

August 18, 2014

Dedication

This work is dedicated to

my two brothers and my sister, each of whom provided inspiration and contributed encouragement throughout, especially my twin brother David, who blazed the trail for the family in earning his PhD, and who has supported me in so many ways spoken and unspoken throughout my life

And to my mother, Caroline, and my father, Elmer, who both endured multiple hardships, but who consistently nurtured my growth and development from my earliest days through their actions, example, and love

And most of all to my wife and children, who supported me and sacrificed for me throughout this work: my children, Talia and Nicholas, who cheerfully accepted (at times) my personal investment in this effort, and especially my wife Amy, without whose support, encouragement, and selflessness none of this would have been possible

Acknowledgments

Many, many people deserve thanks and acknowledgment for their efforts and contributions to this research and my development as a researcher. My collaborators at the University of Michigan both provided access to the dataset of Veterans Health Affairs patients they had compiled over one decade of work, in conjunction with the Veterans Health Affairs Office of Patient Care Services, and provided helpful collaboration during a number of stages of this work. Most valuable was the mentorship I was fortunate to receive from Veterans Health Affairs suicide and pharmacoepidemiological researcher Marcia Valenstein, who has thought more deeply about pharmacoepidemiology than many researches and patiently taught me more than I ever expected to learn in the course of this project. Her acute scientific judgment, pragmatic perspective, personal integrity and dedication to her craft is an ongoing inspiration to me, and I am extremely thankful for her generosity in agreeing to mentor me, a researcher from another state, and give me the benefit of her accumulated knowledge and insight. In addition, programmer Karen Austin dedicatedly performed the programming for the numerous analyses that I specified with great competence and care. Her dedication was particularly impressive given that this work and the approach I adopted stretched her far beyond her usual interests and expertise. While the design and interpretation of every analysis rested with me alone, her thoughtful cooperation contributed substantially to the final product and I am grateful for her willingness to accept such a quantity of detailed instruction from a researcher she had not even met before the initiation of our collaboration. Statistician Myra Kim provided patient statistical guidance and expressed support and encouragement at several important points in the project. John McCarthy also deserves recognition for initiating the compilation of the valuable datasets that were examined in this research and for his support and interest in the project throughout. Others researchers from Ann Arbor such as Amy Bohnert and Amy Kilbourne shared code they had developed that helped specify a few of the propensity score covariates. The entire group was extremely supportive of a researcher from Massachusetts taking a new approach to questions they themselves were interested in.

At UMass Medical School, Jeroan Allison deserves an extreme amount of credit for agreeing to be my Millennium PhD program mentor after meeting with me just once or twice, generously sharing his time over multiple years of the project, never shying away from dedicating however much time it took to make sure I understood a technical topic or an interpersonal point of view. The overall value of his all-encompassing approach to mentorship for me has been truly immense. My thesis chair, Becky

Briesacher generously allowed me to enroll in her class about propensity score methods even before I formally began the Millennium PhD program; her class has taught me more about propensity score methods than any other educational experience I have received. Dr. Briesacher also deserves credit for suggesting adoption of an intent-to-treat approach to the analyses. This simple suggestion opened a whole additional world of questions, interpretation, and implications that may have been completely missed if an "as-treated" perspective – still more of the rule than the exception in pharmacoepidemiological research - had been exclusively adopted in this research. She also helped keep me on track and provided needed encouragement and pragmatism. Other members of my thesis committee also deserve credit: Terry Field, for her ongoing interest and enthusiasm for pharmacoepidemiological research questions and methods, Carl Fulwiler for his thoughtful input as a mental health researcher and clinician, and Mary Jo Pugh, a fellow suicide and pharmacoepidemiological researcher who generously agreed to serve on the committee despite only knowing me through telephone conversations. I am grateful for the time each has provided me in helping me with my PhD and overall research training. Hardy Kornfeld also deserves credit for having the vision to establish the Millennium Ph.D. program, and UMass Medical School for making that vision a reality.

In any complex work, many others also credit, who I will attempt to only name here: David Hoaglin, Kevin Rader, the late Steve Banks, Alexis Henry, Anthony Rothschild, Alan Brookhart, Brian Sauer, Cindy Christiansen, Sue Eisen, Donald Miller, Dan Berlowitz, and Larry Herz. Also making helpful contributions through their encouragement or example was Yolanda Chafee, Kevin Harrison, William Meehan, and C. Arden Pope.

I also would like to thank my wife, children, and brothers and sisters, who both suffered and celebrated through the long process of performing this work. I note my debt to both them and my brothers and sister in the Dedication.

Lastly, I would like to acknowledge my parents, who instilled in me the desire for an opportunity to make a scholarly or intellectual contribution to others, and the perseverance to persist in such efforts. While sadly neither of them was able to live to see this project completed, anything I have accomplished professionally is in large measure due to the nurturing and support they provided me through circumstances that were at best challenging and at worst tragic. I owe them both an extreme debt of gratitude.

Thesis Abstract

Background: The mood stabilizer lithium has long been reported to be associated with reduced suicide risks, but many studies reporting associations between lithium and reduced suicide risks also have been nonrandomized and lacked adjustment for many potential confounders, active controls, uniform follow-up, or intent-to-treat samples. Concerns also have been raised that medications being considered as potential suicide preventative might increase risks of nonsuicide mortality while reducing risks of suicide.

Methods: Three studies of Veterans Health Administration (VHA) patients were conducted combining high-dimensional propensity score matching with intent-to-treat analyses to examine the associations between lithium and valproate and one-year suicide and nonsuicide mortality outcomes.

Results: In intention-to-treat analyses, initiation of lithium, compared to valproate, was associated with increased suicide mortality over 0-365 days among patients with bipolar disorder (Hazard Ratio (HR) 1.50 [95% Confidence Interval 1.05, 2.15]) Nonsuicide mortality among VHA patients with or without bipolar disorder was not significantly associated with the initiation of lithium compared to valproate (HR 0.92 [0.82-1.04]). Rates of treatment discontinuation, however, were very high (≈ 92%). Longitudinal analyses revealed that the increased suicide risks associated with initiating lithium among patients with bipolar disorder occurred exclusively after discontinuation of lithium

treatment. In secondary analyses restricted to patients still receiving their initial treatment, there was no difference in suicide risk between the initiation of lithium or valproate.

Conclusions: Significantly increased risks of suicide were observed at one year among VHA patients with bipolar disorder initiating lithium compared to valproate, related to risks observed after the discontinuation of lithium treatment Since these studies are nonrandomized, confounding may account for some or all of our findings, including the risks observed after lithium discontinuation. Nevertheless, these results suggest that health systems and providers consider steps to minimize any potential lithium discontinuation-associated risk. Approaches might include educating patients about possible risks associated with discontinuation and closely monitoring patients after discontinuation if feasible. Given the obvious importance of any substantive difference between lithium and valproate in suicide or nonsuicide mortality risk, our studies also suggest that further research is needed, especially research that can further minimize the potential for confounding.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIDS/HIV	Charlson Comorbidity Index category designating infection with the human immunodeficiency virus (HIV), often with accompanying Acquired Immunodeficiency Syndrome (AIDS)
ACE Inhibitor(s)	angiotensin-converting-enzyme inhibitor, a class of antihypertensive medications
antidep	antidepressant
ADHD	attention-deficit hyperactivity disorder
AMA	against medical advice
CHF	Charlson Comorbidity Index category designating congestive heart failure
CI	Confidence Interval (specifically, 95% Confidence Interval, and usually designated "95% CI")
Conn	connective, as in "Connective Tissue Disease," a Charlson Comorbidity Index category
CEVD	Charlson Comorbidity Index category designating cerebrovascular disease (most commonly, stroke)
CCM	Charlson Comorbidity Index
COPD	Charlson Comorbidity Index category designating chronic obstructive pulmonary disease
cOR	conditional odds ratio
CPT	current procedural terminology
CT/MRI	CT (computed tomography) or MRI (magnetic resonance imaging) scan
d	days
Dep	dependence, as in the diagnosis "alcohol dependence"
DM	Charlson Comorbidity Index category designating diabetes mellitus, with separate categories indicating that the diabetes is either accompanied with ("w/") or without ("w/o") complications.
DX	diagnosis or diagnoses (alternatively, dx)
Dz	disease
EKG	Electrocardiogram
ER	Emergency Room
GI	gastrointestinal (e.g., "GI illness" indicates an illness of the gastrointestinal tract)
hdPS	high-dimensional propensity score
HepC	Hepatitis C
Hosp	hospital
HR	hazard ratio
HTN	hypertension
ICD-9	International Classification of Disease, 9th Revision
ICD-9-CM	International Classification of Disease, 9th Revision, Clinical Modification
ICU	Intensive Care Unit

ITT	Intention-to-treat, alternatively "intent-to-treat"
JAMA	Journal of the American Medical Association
Li	lithium (alternatively, abbreviated LI)
MAOI	monoamine oxidase inhibitor
MH	mental health
Mod/Sev	Moderate or Severe in severity, as in the Charlson Comorbidity Index category of "Moderate or Severe Liver Disease."
N	sample size (i.e., number of individuals in the sample)
NSAID(s)	nonsteroidal anti-inflammatory medication
NFSB	nonfatal suicide behavior
NFSB, NonMHdx	nonfatal suicidal behavior, diagnosed during a nonmental health hospitalization
NFSB, MHdx	nonfatal suicidal behavior, diagnosed during a mental health hospitalization
NFSB, Outptdx	nonfatal suicidal behavior, diagnosed during an outpatient visit
NonMH	nonmental health
NOS	not otherwise specified
OR	odds ratio
PFT	pulmonary function test
PTSD	post-traumatic stress disorder
PVD	Peripheral Vascular Disease, a category in the Charlson Comorbidity Index
RR	rate ratio
Std. Diff.	standardized difference
SSRI	serotonin-specific reuptake inhibitor
SNRI	serotonin-norepinephrine reuptake inhibitor
TBI	traumatic brain injury
TCA	tricyclic antidepressant
US FDA	United States Food and Drug Administration
U.S.	United States of America
Val	Valproate (alternatively, abbreviated VAL)
VHA	Veterans Health Administration
VISN	Veterans Integrated Service Network, a regional network of VHA hospitals and their associated outpatient clinics
w/	with
w/o	without
yo	years old

CHAPTER I

Introduction

The problem of suicide in the returning military is extremely well-recognized.[1] What particular interventions exist that may be effective for dealing with this important public health issue is far less well-known. In fact, little is known about effective interventions to prevent suicide in any population. For instance, there is only a single medication, clozapine, which has a United States Food and Drug Administration (FDA) indication for the reduction of suicidal behavior. Unfortunately, clozapine is one of the most toxic and closely-regulated medications in all of medicine. (In fact, some commentators have even voiced concerns that the use of clozapine to prevent suicidal behavior might increase mortality from its physical adverse effects approximately as much as it would decrease mortality by reducing suicide [2]). A handful of nonpharmacologic interventions against suicide have also been developed, [3] but, as evidence improves, the effectiveness of some of these interventions also has been called into question. For example, one popular set of public-health oriented interventions, labeled "means restriction" has been found effective in some studies [4] but not others.[5] Finally, many of the nonpharmacologic interventions provided in clinical settings (e.g., particular psychotherapies[6, 7]) are also highly resource-intensive, requiring specialized training and sophisticated implementation.

In addition to that paucity of rigorous evidence concerning pharmacological and non-pharmacological interventions to address suicide risk, concerns have recently increased that several psychiatric medications intended to treat psychiatric conditions (antidepressants and anticonvulsants),[8, 9] may actually increase, rather than decrease, the risk of suicidal ideation or behavior.

***

Among the candidate pharmacological interventions currently available that lack FDA indications to reduce suicide or suicidal behavior, the medication that is best supported by existing evidence is lithium. Ever since a well-publicized meta-analysis (primarily of observational studies) was completed 15 years ago reporting that lithium appeared to be associated with uniquely large reduction in suicide risk (5.6 to 8.6-fold reduction in risk),[10] many psychiatrists have been aware of the possibility that lithium treatment might reduce suicidal behavior and death by suicide. Treatment guidelines even went so far as to highlight this possibility.[11] Subsequently, in data less appreciated by many psychiatrists, a meta-analysis of limited randomized evidence also observed a significant association of lithium treatment with reduced suicidal behavior and death by suicide.[12] However, these studies in general have not focused on Veterans, and this evidence base, while encouraging, is beset with a number of limitations.

One of the biggest concerns for the meta-analysis including observational findings is the conceivable (and perhaps even probable) possibility that until very recently patients started on lithium treatment might have been at lower risk of suicide than comparator medications, given that lithium may in certain circumstances increase risks of self-harm

upon overdose.[13, 14] Thus, this meta-analysis, since it includes a number of nonrandomized studies from early in the use of lithium, likely summarizes study findings that are biased in favor of finding protective associations for lithium. In keeping with this possibility, effect sizes have declined with subsequent updating of this meta-analyses.[15]

This meta-analysis helped inspire a large, well-publicized cohort study in 2003 published in the *Journal of the American Medical Association (JAMA)*.[16] This study found lithium to be associated with significantly lower risk of suicide than valproate, with a large effect size (HR = 0.37, 95% CI 0.16, 0.91, p=0.03). Lesser but still significant associations were observed for lithium compared to valproate for reduced suicide attempts requiring hospitalization (HR=0.59, 95% CI 0.43-0.83, p=0.002) or emergency room visits (HR=0.56, 95% CI 0.45-0.71, p<.001).

This study, however, also contains the potential for numerous possible biases, including poorly-described covariates, a lack of information concerning the initial imbalances in many of these covariates among treatment groups, and, potentially most importantly, longer follow-up times for the lithium compared to the valproate treatment arm.[16] This substantial difference in follow-up times likely resulted in a larger proportion of time at-risk for the lithium cohort accruing from the later stages of maintenance treatment. Later stages of treatment, both through selection occurring during treatment and through the increasing passage of time from initial medication initiation (frequently observed to be a high-risk period for suicide or suicidal behavior[17, 18]) is likely to be a period of considerably lower suicide risk. Permitting the lithium treatment arm to preferentially accrue longer follow-up time provides a comparison that is likely to be to

some degree biased. One additional important aspect of *JAMA* cohort study was the observation, noted only extremely briefly, that suicide risk shortly after discontinuation of study medications was particularly pronounced.[16] This is a concern because of literature suggesting prominent suicide and suicidal behavior risks occur shortly after lithium discontinuation,[19-21] however it is unclear whether, while prominent, these risks are distinctive to lithium or are also associated with valproate discontinuation as well.[22, 23]

Subsequent to the (primarily) nonrandomized study meta-analysis and the *JAMA* cohort study, a meta-analysis restricted to randomized clinical trials was published in 2005.[24] (This meta-analysis was subsequently revised in 2013,[25] but the results are less succinctly summarized since no overall comparison of lithium to other active comparators is provided). The original meta-analysis (2005) observed that lithium was significantly associated with a substantial reduction in suicide risk (HR = 0.26, 95% CI 0.09, 0.77), although only 9 suicides were observed, including only 2 among patients receiving lithium.

Most recently, a small randomized trial (n=49 patients randomized per treatment) designed specifically to examine the comparative effectiveness of lithium and valproate for suicidal behavior observed no significant different between the treatments, and only modest nonsignificant effect sizes (the time to first episode of suicidal behavior, or concerns about potential suicidal behavior leading to a change in treatment, were approximately 16-25% lower in the lithium treatment arm).[26] The 2013 meta-analysis combines these results with secondary findings from a recent trial examining the effectiveness of lithium and valproate in preventing mood episodes[27] to give an overall

estimate for the difference in suicidal behavior risk between lithium and valproate treatment of OR = 0.64 (95% CI 0.30, 1.36, p=0.24).[25]

***

One additional aspect of studies of suicide and suicidal behavior risk involves the concerns that have been raised that treatment studies of suicidal behavior may mischaracterize, to some degree, associations between treatments and suicide risk. This is suggested in the pivotal trial observing efficacy for clozapine focused in preventing repeat suicidal behavior.[28] While significant reductions in suicidal behavior was observed for patients treated with clozapine compared to olanzapine, doubt still exist about the overall effectiveness of this medicine to treat the most serious consequences of suicide behavior, death by suicide. The reduction in risk of suicidal behavior did not appear to straightforwardly generalize to suicide, although numbers of suicides were very small (5 suicides were observed among patients initiating clozapine versus 3 suicides among patients initiating olanzapine).[28]

Adding to this concern is the fact that generally only about half of suicides occur among individuals with prior suicidal behavior; this implies for half of suicide deaths the "pathway" to suicide death does not even include suicidal behavior. Finally, certain risk factors for suicidal behavior have an opposite association for suicide risk. Regarding gender, 3-4X as many women as men attempt suicide, but death by suicide is 3-4X as common in men versus women. Similarly, with respect to age, suicide attempts are the most common in young people, but suicide deaths are most common in the elderly. Even for lithium, some meta-analyses have observed that lithium appears to shift (decrease) the

ratio of suicides to suicide attempts, implying a greater association with reduced suicide than suicide behavior risks.[29] Thus, concerns exist that investigating the more numerous (and convenient, for purposes of demonstrating statistical significance) events of suicidal behavior may convey an incomplete or potentially misleading picture of what interventions are the most likely to prevent death by suicide.

One final reason supports the investigation of death by suicide rather than, or in addition, to suicide attempts: the completeness of outcome data. Events of suicidal behavior may fail to be documented in Veterans Health Administration (VHA) or any other non-universal health system records due to the receipt of out-of-system emergency care. In contrast, deaths (by suicide or any other cause) are comprehensively recorded nationwide in the National Death Index, the accepted "gold standard" for mortality studies.

***

The debates about clozapine mentioned above have highlighted another important issue in evaluating the ultimate clinical utility of any candidate suicide preventative: the need to ensure that any intervention designed to prevent a relatively infrequent source of mortality, suicide, does not create risks for other causes of mortality that could easily outweigh any benefits in suicide risk reduction. While suicide is a highly important, tragic, and potentially preventable cause of mortality (suicide currently ranks as the 10th leading cause of death in the United States), the incidence of suicide is much less frequent, on a population basis, than some other causes of death (most notably cardiovascular mortality). Thus, comparatively small associations with increased

cardiovascular mortality associated with a candidate suicide preventative could easily more than counterbalance benefits resulting from even a substantial association with decreased suicide risk.

Furthermore, the question of the potential associations of lithium (and comparison psychiatric medications) has its own independent scientific relevance. Lithium is a very small ion (an alkali metal that is analogue to sodium and potassium) that penetrates almost all body tissues and has long been recognized to have potent biological effects. As early as 1892, experiments indicated that lithium can have profound influence on embryo development in sea urchins, and now it is recognized to influence development in a variety of organisms such as frogs and slime molds.[30] It was later determined that lithium influences key second messenger systems such as phosphoinositides, and more recent investigations have found that lithium affects a host of other cellular targets as well.[31] These cellular impacts likely lead to lithium's well-known diverse impacts on physiology at the organ level. These physiological effects range from the potentially beneficial (e.g., leukocytosis,[32] reduced heart rate,[33] and neurogenesis[34]) to the potentially hazardous (e.g., renal[35, 36] and thyroid insufficiency,[37] QTc prolongation,[33] or arrhythmias[38-40]).

Trials designed to compare psychiatric interventions with a primary focus on suicide mortality, or even nonsuicide mortality, however, are exceedingly unlikely to occur. First, mortality is not generally appreciated or thought about as a potential consequence of psychiatric treatment. Thus it is unlikely that any trial focusing on such outcomes, at least as a primary outcome, will be acceptable to patient volunteers.

Moreover, many instances of poor outcomes in psychiatry (e.g., suicides or suicidal behavior) are preceded by clinical deterioration among patients (increased psychiatric symptoms, deteriorating psychosocial circumstances, and the initiation or worsening of substance abuse). This potential foreshadowing raises the ethical responsibility during the trial to intervene (e.g., adding further medications, psychiatric hospitalization) to prevent the very outcome being studied. In fact, modern clinical trials of suicidal behavior risks have used as the primary outcome both hospitalizations and/or other interventions (such as discontinuation, addition to, or change of the intervention medication) as part of a composite outcome. Although highly necessary and desirable from an ethical and humane standpoint, it is well-recognized that these composite outcomes likely have a lack of comparability, to some important degree, with actual suicidal behavior or death by suicide. The correlation of these composite outcomes with suicidal behavior or particularly death by suicide is particularly uncertain.

Thus, investigating the association of lithium and comparison medications of greatest public health interest on the outcomes of greatest public health interest (death by suicide and from all other causes), a major element of any evidence base is likely to be rigorously-conducted, extremely large-scale observational studies. The next section describes some of the important challenges and opportunities of such research and the major approaches adopted in these studies to address these issues.

***

Although advances in nonrandomized comparative effectiveness research have been burgeoning over the past few years, fueled in part by major funding and

organizational initiatives focusing on this important field of health care research, it is well-known that nonrandomized studies of interventions have a greater potential for bias than randomized studies, and even than nonrandomized studies which are not investigating interventions. This is due to the difficult-to-address phenomenon of confounding by indication. In general, important clinical decisions made by providers and patients are likely to be influenced by factors that may potentially impact risk for outcome, and thus potentially bias the results of studies. This poses a dual challenge. The first is, rather than address just a few pertinent risk factors (e.g., age, gender, race), as might be done in a prognostic or descriptive epidemiological study, a host of potentially clinically relevant risk factors may need to be examined, adjusted for, or controlled to ensure that confounding by indication is not overly biasing results. These additional factors for consideration include, but are not limited to indicators of the severity the patient's diagnoses/conditions, as reflected by the presence or absence of recent hospitalizations, current and recent medications taken for those conditions, the numbers and types of medical or psychiatric specialists recently seen, recent diagnostic tests received, as well as indicators of general mental and nonmental health care utilization and other demographics (e.g., rural versus urban residence, marital status, etc.).

Among recent analytic innovations to attempt to address confounding by indication, an recently proposed approach likely to be particularly suited for application to rich observational datasets is the high-dimensional propensity score. Introduced in 2009, this method attempts to more fully address potential confounding by substantially

increasing the number of covariates that conceivably can be included in a model. Many analytic methods are limited in the number of covariates that can be included through the number of outcome events. However, propensity scores, by modeling the influence of outcome-associated measured covariates on exposure, rather than outcome, provide a mechanism to control for 10X or even 100X more covariates than past approaches. High-dimensional propensity scores formalize this approach by deliberately attempting to have this more numerous collection of covariates incorporate a broader range (i.e., dimensions) of potential confounding as well. The initial manuscript describing this method included a number of covariates in each of 5 covariate dimensions. While most analyses were restricted to 200 to 500 covariates, already a large number, one comparison in this initial manuscript included 4989 covariates in the model.[41]

In addition, this technique gains added power when combined with matching or weighting methods. These methods, especially propensity score matching, provide closer balance in the included covariates than methods that stratify on a limited number of categories (e.g., quintiles) [42] or incorporate the propensity score as a regression covariate. In addition, matching and weighting also have the beneficial property that these methods mimic in some ways the control of covariates obtained in a randomized trial. The balance in pertinent covariates can be presented and readily grasped in a "Table 1" describing the treatment groups. A very important limitation is, however, in contrast to a randomized trial, this balance only pertains to the measured covariates (and is sensitive to considerations of how completely they are measured and/or how correctly they are

modeled), and not to unmeasured covariates. (Successfully randomized trials balance both measured and unmeasured covariates).

The development and use of high-dimensional propensity scores has not been without controversy, especially concerning how variables to be included are selected. In general, the high-dimensional propensity score leans in the direction of data-driven, rather than theory-driven, criteria for inclusion of covariates. While this may seem attractive to some, these data-driven efforts have included tendencies to weight imbalances in prevalence between the treatment arms in variable selection (rather than simply prioritizing covariates based on their association with outcomes, rather than treatment exposure). This approach of selecting covariates for inclusion based on part with their association with treatment exposure may introduce biases in the estimate of outcome risks. In fact, it has recently been recognized that such a variable selection strategy has the potential to enhance a potential limitation distinct to propensity scores, the phenomenon of "*bias amplification"* (or the term we prefer, confounding amplification).[43-47] Confounding amplification refers to the tendency as a propensity score becomes progressively more predictive of treatment exposure for the method to actually enhance, or increase, differences in unincluded confounders. Stated another way, while the intent of high-dimensional propensity scores are to incorporate greater and greater confounding information into the score, leaving less confounding "unincluded," at the same time the propensity score will be operating to enhance whatever small or large residual confounding after the large number of included covariates are brought into balance. This potential inherent limitation to propensity scores may be especially

relevant to studies in which the likelihood of some unmeasured confounding is high, such as studies of mental health interventions.

***

In addition to investigating the application and findings resulting from application of high-dimensional propensity scores to the rich VA clinical databases, the studies described here also incorporate several other important methodological enhancements compared to previous studies. The most important of these enhancements include the quantification of intent-to-treat and "former user" risks. Intent-to-treat estimates, still not routinely used for nonrandomized studies, perform two extremely valuable functions: 1) they help make effect estimates from nonrandomized studies more directly comparable to findings from randomized trials, and 2) they ensure that any risks arising after treatment discontinuation are considered in judgments of treatment effectiveness. (As indicated above, previous research is suggestive that such risks may exist). Risks observed among "former users" in the survival analyses refers to the risk observed for some patients after they discontinue their initiated treatment until they resume either treatment. These risks, observed during a period of non-exposure to treatment, can serve as a useful index suggesting the amount of residual baseline confounding or confounding arising from selection of patients to discontinue treatment.[48, 49]

An additional valuable facet to the studies described below is that temporal changes in confounding by indication, coupled with the restriction of our study to a more recent time period than most prior literature (1999-2008) is expected to provide a useful contrast to prior literature in an analytic sense. Since our studies were restricted to the

period subsequent to or coincident with the publication of the meta-analyses, cohort studies, and treatment guidelines supporting lithium's potential effectiveness versus suicide and suicidal behavior, the pattern of confounding may be substantially different in this study from much of the previous evidence. (Even the 2003 cohort study was restricted to the period 1995 to 1999, before this literature had emerged and during a period in which valproate use might have been reserved for those who were most treatment-refractory). While confounding by indication is never advantageous, the fact that confounding by indication may have reversed from prior periods so that higher risk patients were preferentially has some advantages. If a comparable beneficial association between lithium and suicide risk can be observed comparable to what has been observed in past studies, then confidence would increase that at least some of this association is genuine may be increased given that the direction of bias deemed likely would have biased against observing this association, or at least favored this association less than previous literature.

There are other more basic aspects to study design that may have methodological advantages compared to prior studies, such as the use of incident-user samples,[50] and modelling of potential baseline confounders in flexible forms (i.e., approaches which do not assume linearity) that also allow risks to vary substantial over time in the period preceding treatment initiation. Also of importance is the focus on a fixed, uniform follow-up period after treatment initiation of likely clinical relevance: the first year after initiation. The first six months to one year after treatment initiation is likely to be the focus of any clinical intervention particularly designed to address short-term suicide risk.

Finally, the studies reported here also have the methodological advantage of examining lithium use in two contexts: one in which confounding by indication may be highly prominent (suicide), and a second (nonsuicide mortality) where at least direct, conscious confounding by indication engendered by prescriber behavior would be expected to be more limited or potentially even minimal.

This set of studies sought to apply the techniques of high-dimensional propensity scores, intent-to-treat estimates, and related techniques, to the crucial question of preventing suicide in Veterans. These studies include a preliminary examination of the larger question of determining whether any suicide prevention benefits of a candidate intervention such as lithium are outweighed by mortality risks from other causes. In the chapters that follow, we describe the implementation and findings of three studies designed to evaluate whether lithium is associated with reduced risks for suicide and to further investigate its association with other causes of mortality. These studies are intended to both incorporate a sophisticated awareness of the potential limitations of nonrandomized treatment studies while advancing the evidence base concerning whether lithium should be considered as a candidate intervention to reduce suicide.

CHAPTER II

Suicide Risk in Veterans Health Administration Patients with Mental Health Diagnoses Initiating Lithium or Valproate

Abstract

Background: Lithium has been reported in some, but not all, studies to be associated with reduced suicide risk.

Methods: Intention-to-treat, high-dimensional propensity score-matched cohort study of one-year suicide mortality in Veterans Health Administration patients (n = 21,194/treatment) initiating lithium or valproate from 1999-2008. One-year suicide mortality was examined among all patients initiating lithium or valproate (intention-to-treat analysis) who were matched on 934 propensity score covariates ranging from demographics, diagnoses, inpatient and outpatient encounters, current and recent medications, and indicators of recent suicidal behavior.

Results: No significant difference in suicide was observed over 0-365 days in the primary intent-to-treat analysis (lithium/valproate conditional odds ratio (cOR) = 1.22, 95% CI 0.82, 1.81; p = 0.32), when individuals were still receiving their initial lithium or valproate treatment (cOR = 0.86, 0.46, 1.61; p = 0.63) or after such treatment had been discontinued/modified (OR = 1.51, 95% CI 0.91, 2.50; p = 0.11). Significantly increased suicide risks in the first 0-180 days of follow-up were observed in secondary analyses of

individuals after discontinuation of their initial lithium, compared to valproate, treatment (OR = 2.72, 95% CI 1.21, 6.11; p = 0.015).

Conclusions: We detect no significant differences in one- year suicide risk between patients initiated on lithium compared to valproate. However, we did observe an increased suicide rate shortly after individuals discontinued initial lithium compared to valproate, treatment. The implications of this secondary finding would differ depending on whether the increased risk upon discontinuation relates more to risks associated with discontinuation of lithium versus valproate, or to the characteristics of the patients initiating and subsequently stopping each treatment (confounding). Further research is needed to determine initiating lithium therapy may reduce suicide risk during active treatment, increase risk upon discontinuation, or both. In the meantime, these results suggest that health systems and providers should consider approaches to minimize any potential lithium discontinuation-associated risk, including educating patients about possible risks associated with discontinuation and closely monitoring patients after discontinuation.

Introduction

Reducing suicide is both an international priority and a particular need for Veterans.[1] The mood stabilizer lithium has been reported to be associated with uniquely large reductions in risks of suicide and suicidal behavior.[15, 16, 29] Many studies, however, have been nonrandomized and contained substantial methodological limitations, lacking adjustment for many potential confounders, active controls, incident-user designs, uniform follow-up, or intent-to-treat samples.[14] A meta-analysis of 32 randomized trials of lithium reported significant reductions of suicide risk with lithium,[12] as did a recent trial of lithium augmentation,[51] but these findings were based on an extremely small number of outcomes (i.e., only 2 suicides total among patients receiving lithium).

In contrast, other nonrandomized studies[22, 52] and a recent small randomized trial of suicidal behavior[53] report smaller, nonsignificant reductions in suicide or nonfatal suicidal behavior associated with lithium in comparison to another commonly-used mood stabilizer, valproate. While nonrandomized studies provide the large sample sizes needed to determine associations between treatments and suicide, concerns exist that prior nonrandomized studies may have been confounded through preferential prescription of lithium to patients at low[13] or lower[14] risk of suicide. Prescriber behavior may have changed substantially over the past 15 years, however, given well-publicized meta-analyses,[10, 20] treatment guidelines,[11] and high-profile studies[16] reporting that lithium treatment may be associated with distinct reductions in suicide risk.

Using data from this more recent period, we sought to address confounding through methods intended to approximate some of the strengths of randomized trials. In

addition to adopting a fixed follow-up time, these methods included: 1) matching patients based on a "high-dimensional" propensity score[42] designed to be particular comprehensive in including information on potential confounders available from VHA administrative and clinical databases,[41] and 2) deriving intent-to-treat and post-discontinuation risk estimates.[48] Propensity score matching can permit an unusually large number of covariates to be controlled, creating treatment groups closely similar in prevalence for numerous covariates (similar to a trial). Intent-to-treat estimates ensure that confounding arising after treatment initiation from difference in the patients discontinuing each treatment to be controlled, and also have the advantage of ensuring that risks arising after treatment discontinuation are considered in judgments of treatment effectiveness. Combining data from this more recent epoch of care with these methodological approaches, we conducted the largest cohort study to date examining whether suicide risk differs between patients initiated on lithium compared to valproate.

Methods

Data Sources

Demographic, inpatient and outpatient mental and nonmental health treatment records, and outpatient pharmacy prescription data was obtained from the Veterans Health Administration (VHA) National Psychosis and Depression Registries[54] (linked, de-identified healthcare databases of all VHA patients since 1997 with at least one psychotic or depressive disorder diagnosis). This study was approved by the Institutional Review Boards of the Bedford and Ann Arbor Veterans Affairs Medical Centers.

Study Cohort

Incident users (≥ 6 months of no lithium or valproate use) with recent VHA utilization (past year and a previous year) were identified among all patients with mood or psychotic disorders within the past 30 days receiving at least one outpatient prescription for lithium or valproate from April 1999 to December 2008. These broad diagnostic inclusion criteria (Appendix 1-1) were chosen to maximize statistical power, given the comparatively few suicides expected, even in a large cohort, over a fixed one-year follow-up period, and to facilitate the evaluation of lithium and valproate as widely-useful suicide preventatives. Prior research suggests that any effectiveness of lithium against suicide is not restricted to patients with bipolar disorder.[51, 55] At least 89% of our cohort of the ultimate propensity score-matched cohort had an affective disorder (Appendix 1-1). Individuals with potentially nonpsychiatric indications for treatment were excluded (epilepsy, cluster or migraine headache, or neuropathy diagnoses in the past 30 days; dementia medication use in the past 180 days; cancer, dementia, skull fracture diagnosis, traumatic brain injury diagnosis or treatment; home care or hospice care in the past year; or any nursing home residence or inpatient rehabilitation in the past 2 years). Patients were also excluded if they had started their mood stabilizer on an "as needed" basis or both mood stabilizers simultaneously (Appendix Figure 1-1).

Exposure Determination

Receipt of lithium or valproate was determined by outpatient prescription fills. For the primary "intent-to-treat" analysis, all individuals initiating treatment were

followed until end of follow-up (365 days), suicide, or death from other causes. Secondary analyses examined briefer follow-up times and/or stratified follow-up time by whether individuals were still receiving initial treatment. Individuals were identified as "still receiving initial treatment" if they had not switched to or added the other treatment, nor discontinued the initial treatment (a ≥ 15-day gap between outpatient prescriptions, adjusted for early refills). All other follow-up time was classified as occurring during the period when individuals had "stopped/modified" initial treatment. Since this stopper/modifier group included individuals subsequently resuming either treatment, we secondarily analyzed suicide risks for individuals over 0-180 days that had stopped initial treatment and not resumed either treatment before suicide, other mortality, or the end of follow-up. Risks observed after treatment discontinuation may reflect risks related to discontinuation of the medication (e.g., "rebound" mania or depression), but may also reflect differences in baseline risk between the treatment groups (i.e., confounding) still remaining in the analysis, or differences in selection occurring during follow-up.[48]

Outcome

Date and cause of death (suicide) was obtained from National Death Index files[56] for 1999-2009 using previously established definitions (International Classification of Diseases, Tenth Revision codes X60–X84, Y87.0, and U03).[57]

Propensity Score Modeling

An extensive set of 934 baseline covariates was derived (Table 2-1 and Appendix 1-2) from VHA databases reflecting demographics, diagnoses, general VHA mental and

nonmental health healthcare utilization,[41] hospitalizations, clinic use, diagnostic testing, current and recent prescriptions (Appendix 1-3), diagnosed suicide attempts and injuries, regional (state-level) suicide risk, and prior mood stabilizer treatment, often modeled over several time periods or with flexible forms (multiple indicator variables). This approach follows the general aims of "high-dimensional" propensity score methodology,[41, 58] but did not include automated variable generation or selection. Instead, a number of covariates in each utilization/diagnosis/medication domain were generally included, since the full determinants of suicide risk are not well understood, although covariates with a substantial association with treatment exposure were individually evaluated and removed if they were judged unlikely to be confounders. (Inclusion of variables that are substantially associated with exposure but are not confounders can actually increase confounding from uncontrolled factors).[43, 44, 46, 59]

Statistical Methods

The propensity score was calculated by logistic regression (c statistic=0.69). We then 1:1 matched patients initiating lithium and valproate using propensity score calipers of 0.03 (0.2 standard deviations of the propensity score logit)[60, 61] achieving 98.7% matching of lithium-initiated patients. Balance in the prevalence of covariates between

Table 2-1. Summary of Variables Included in the Propensity Score
(Prevalence of each balanced within a standardized difference of <0.018 in final matched cohort)

Type of Patient Characteristic	Covariates Included
	General Covariates
Demographics	49 Total Covariates including: Age (5-year categories), Sex, self-reported Race, Ethnicity, Marital Status, Income, Disability Status, Distance to VHA facility, Urban/Rural hospital location, and Year of Medication Start
	Psychiatric Covariates
Presenting Diagnosis	9 Variables, including Bipolar I, Bipolar II, Bipolar NOS, Major Depression, Depression NOS, Schizophrenia, Schizoaffective Disorder and Other Psychoses
General Utilization	74 Covariates, including Total VA Mental Health (MH) Provider Visits x 6 time periods, Total MH hospitalizations x 2 time periods, Total Current MH Medications, Recently Discontinued MH medications, and Possibly Discontinued MH medications, and Total Diagnostic Interviews, Total Medication Management Visits, Total Individual Psychotherapy Visits, Total Group Psychotherapy Visits (all x 2 time periods)
Comorbid Diagnoses	46 Covariates including PTSD, Other Anxiety Disorders, Adjustment Disorders, Personality Disorders, Somatoform Disorders, Impulse Control Disorders, Sleep Disorders, Eating Disorders, Sexual Disorders, Delusional Disorder, ADHD, Development Disorders, Cognitive Disorder NOS, and Dissociative Disorder

Table 2-1. (continued)

Comorbid Substance Abuse Diagnoses	41 Covariates, including 4 Covariates (Abuse, Dependence, Remission from Abuse, and Remission from Dependence for each of the following: Alcohol, Amphetamines, Cocaine, Marijuana, Opioids, Sedatives, Other), with other covariates for Hallucinogen Abuse/Dependence/Remission, Combined Drug Dependence and Remission from Combined Drug Dependence, with and without opioids, Unspecified Dependence, and Alcohol intoxication and Alcohol or Drug psychoses
Suicide Attempt Diagnoses	9 Covariates, designating if Suicide Attempt diagnosed during NonMental Health (NonMH) hospitalization, MH hospitalizations, or as Outpatients, x 3 time periods (0-30d, 31-180d, 181-365d).
Hospitalizations	10 Covariates, including whether Inpatient on Start Date, Within last 7 days, 8-30 days, 31-180 days, and 181-365 days, Type of Latest Hospitalization (Psychiatric, Substance Abuse, Residential/Day program, Domiciliary), and whether any hospitalizations in last year involved an AMA discharge
Specific Outpatient Utilization	48 Covariates (all modeled as 0 visits, 1 visits, or 2+ visits): General Mental Health clinic, Psychiatry visits, Psychotherapy visits, Substance Use Disorder visit, Primary Care Mental Health clinic, Health Care for Homeless Veterans, and Substance Abuse and non-Substance Abuse visits, x 2 time periods (0-180 days and 181-365 days)
Current Medications	24 Covariates, including Olanzapine, Risperidone, Quetiapine, Ziprasidone, Aripiprazole, Clozapine, First Generation Antipsychotics, Other Mood Stabilizers, SSRIs, SNRIs, Bupropion, Mirtazapine, TCAs, MAOIs, Benzodiazepines, other Hypnotics, Stimulants, Substance Abuse treatments, and other medications
Recent Medications	24 covariates, designating prescription received in last 180 days for same medication categories as “Recent Medications” but no prescribed supply extending to Lithium/Valproate start date

Table 2-1. (continued)

Prior Treatment History	Prior treatment with Any Mood Stabilizer, Prior treatment with Lithium or Valproate
Geographic Suicide Risk	5 variables designating quintiles of Age-Adjusted State Suicide Rates (2000-2007)
NonPsychiatric Variables of Possible Particular Relevance to Suicide Risk	
Nonpsychiatric Diagnoses	7 covariates, including any Acute Injury, any Fracture, Blood Vessel injury, Internal injuries, Open Wounds, Poisoning, Inhalation/Drowning/Asphyxiation injury
Nonpsychiatric Utilization	6 covariates, including Pain Clinic visits (0, 1, 2+) x 2 time periods
Nonpsychiatric Medications	4 variables, including current Opiate Pain Medicine, recent Opiate Pain Medicine, and 2 types of overdose antidotes
Also Included:	
Numerous covariates designating Nonpsychiatric Diagnoses, Nonpsychiatric Hospitalizations, Nonpsychiatric VHA Utilization (General and Specific), Nonpsychiatric Medications (current and recent), and Nonpsychiatric Diagnostic Tests. 3 covariates recording prior VHA pharmacy use (any prior use, use in the last 180 days, and use in the last 365 days) were also included to help balance the extensiveness of pharmacy records among recipients.	

the treatment groups were assessed using standardized differences (Table 2). Standardized differences are equivalent to Cohen's *d* effect sizes, with a difference of ≥ 0.10 often considered as indicating significant imbalance.[42]

For the analyses of the intent-to-treat cohort or of individuals still receiving initial treatment, we used conditional logistic regression, whereas for individuals stopping or modifying treatment, ordinary logistic regression was used since matching was not rigorously preserved.

Several additional secondary analyses were conducted, including comparing the prevalence between the treatment groups of diagnostically-coded suicidal ideation (V62.84, a code for suicidal ideation only available for the years 2005-2008) in the 30 days prior to treatment initiation among the patients for whom this information was available (< 50% of the full sample). We also compared the suicide risk associated with the treatment groups prior to matching, conducted a Cox regression analysis, and conducted a sensitivity analysis matching the sample with an alternative propensity score. All analyses were performed using SAS, version 9.3, except the standardized difference calculations (Microsoft Excel 2007).

Results

A 1:1 propensity-score matched cohort of 42,388 patients (including 102 suicides over 365 days of follow-up) was derived from 93,335 incident users of lithium or valproate meeting inclusion/exclusion criteria (Appendix Figure 1). Patients initiating lithium and valproate were generally balanced even prior to matching on a wide variety of psychiatric and nonpsychiatric diagnoses, outpatient and inpatient utilization, and

medication covariates: only 17 (1.8%) of 934 covariates exhibited initial standardized differences of ≥0.10 between treatment groups. Table 2-2 demonstrates the close balance achieved after propensity score matching between treatment groups for these 17 initially-imbalanced covariates plus 14 established suicide risk factors.[62-66] Similarly close balance after matching was observed for all other covariates (none of the 934 covariates had a standardized difference of even 0.018 after matching). Despite the general balance in most covariates observed between the treatment groups prior to matching, propensity score matching led to a substantial reduction in the observed treatment effect estimate (initial 0-365 day Odds Ratio [OR]=1.45 [lithium/valproate], versus 0-365 day Conditional Odds Ratio (cOR)=1.22 after propensity score matching).

The treatment groups also displayed very substantial, but highly similar, rates of stopping or modifying initial treatment: 47% of patients initiated on both lithium and valproate were still receiving their initial treatment at 90 days, 24% at 180 days, and only 8% at 365 days (Appendix Table 1-1).

Table 2-3A and 2-3B provides results for these extensively-matched treatment groups over the first year after medication initiation. No significant difference was noted in the primary outcome (intent-to-treat 0-365 day cOR 1.22, 95% Confidence Interval [CI] 0.82, 1.81; p=0.32). In addition, no significant difference was noted in a secondary analysis among patients during the period within the first year in which they were still receiving initial treatment (cOR 0.86, 95% CI 0.46, 1.61; p=0.63), an effectiveness

Table 2-2. Characteristics of Patients Initiating Lithium and Valproate (Propensity-score Matched Sample)

Patient Characteristic[a]	Lithium (n=21194) N	(%)	Valproate (n=21194) N	(%)	Standardized Difference
Demographics					
Age ≥50 years old[b]	10244	48.3	10156	47.9	0.008
SEX (Female)[c]	2894	13.7	2934	13.8	0.005
RACE, White	16748	79.0	16793	79.2	0.005
RACE, Black	2825	13.3	2770	13.1	0.008
Married	7416	35.0	7298	34.4	0.012
STATE SUICIDE RATE, 3rd quintile	3305	15.6	3251	15.3	0.007
Presenting Diagnosis[d] (Past 30 days)					
BIPOLAR I DISORDER	9562	45.1	9683	45.7	0.011
BIPOLAR DISORDER, Not Otherwise Specified (NOS)	1643	7.8	1661	7.8	0.003
DEPRESSIVE DISORDER, Not Otherwise Specified (NOS)	4214	19.9	4129	19.5	0.010
SCHIZOPHRENIA	924	4.4	949	4.5	0.006
OTHER PSYCHOSIS	252	1.2	255	1.2	0.001
Additional Psychiatric Diagnoses (Past Year)					
POST-TRAUMATIC STRESS DISORDER (PTSD)	4842	22.8	4749	22.4	0.010
Alcohol Dependence	4426	20.9	4478	21.1	0.006
Recent Suicidal Behavior Diagnoses (past 30d, by location where diagnosed)					
NonMental Health Hospital-Diagnosed	28	0.13	24	0.11	0.005
Mental Health Hospital-Diagnosed	30	0.14	32	0.15	0.002
Outpatient Visit-Diagnosed	144	0.68	147	0.69	0.002
Recent Suicidal Behavior Diagnoses (past 31-180d, by location where diagnosed)					
NonMental Health Hospital-Diagnosed	43	0.20	43	0.20	0.000
Mental Health Hospital-Diagnosed	31	0.15	29	0.14	0.003
Outpatient Visit-Diagnosed	90	0.42	82	0.39	0.006
Possible Suicidal Behavior-Related Diagnoses (past year)					
Any Acute Injury	3872	18.3	3884	18.3	0.001
Recent Discharge from Psychiatric Hospitalization					
Discharged in past 7 days	2232	10.5	2219	10.5	0.002
Discharged in past 8-30 days	863	4.1	881	4.2	0.004
Discharged in past 31-180 days	2024	9.5	2063	9.7	0.006

Table 2-2. (continued)

Current Psychiatric Medications					
OTHER MOOD STABILIZERS(S)	2891	13.6	2854	13.5	0.005
SSRI antidepressant	7615	35.9	7666	36.2	0.005
SNRI antidepressant	1988	9.4	2019	9.5	0.005
Past Treatment History					
PRIOR MOOD STABILIZER	7503	35.4	7530	35.5	0.003
Diagnoses, Nonpsychiatric (past year)					
MILD LIVER DISEASE	1747	8.2	1719	8.1	0.005
Outpatient Utilization, Nonpsychiatric (past 180d)					
GASTROENTEROLOGY CLINIC, 1+ visits	1102	5.2	1077	5.1	0.005
Current Medications, Nonpsychiatric					
THIAZIDE DIURETIC	1499	7.1	1492	7.0	0.001
ACE INHIBITOR	2764	13.0	2736	12.9	0.004
NSAIDS	3491	16.5	3522	16.6	0.004

[a] Patient Characteristics listed in ALL CAPITAL LETTERS are the 17 characteristics with a ≥ 0.10 Initial Standardized Difference (although these differences are far less after the propensity-score matching).

[b] Age presented in this format (<50 years vs. ≥50 years old) to streamline its presentation within this Table. Age was actually modeled using 11 indicator variables reflecting age groups from < 35 years old in 5-year intervals to ≥80 years old.

[c] The proportion of females in the cohort is low because the Veteran sample is predominantly male.

[d] Percentages for Indicating Diagnoses do not add up to 100%. Some diagnoses were not substantially imbalanced and therefore not included in the Table, although they were included in the propensity score and matched upon (e.g. Major Depression, Bipolar II Disorder, Schizoaffective Disorder, and ≥2 Indicating Diagnoses in past 30 days).

COVARIATES in ALL CAPITAL LETTERS designate those with an initial imbalance ≥ 0.10 standardized difference prior to matching

Table 2-3. Suicide Deaths and Rates over Time by Mood Stabilizer Treatment

Table 2-3A. Primary Analysis (Intent-to-treat, 0-365 days)

	All Patients Initiating Treatment (Intent-to-Treat Cohort)							
	Patients Initiating Lithium			Patients Initiating Valproate				
Follow-up Time	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Conditional Odds Ratio (95% CI)	Rate Ratio
0-365 days	21194	56	7.27	21194	46	5.98	1.22[a] (0.82-1.81)	1.22

Table 2-3B. 0-365 day Findings Stratified by Initial Treatment Status

	During Exposure to Initial Treatment[b]							
	Patients Initiating Lithium			Patients Initiating Valproate				
Follow-up Time	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Conditional Odds Ratio (95% CI)	Rate Ratio
0-365 days	21194	18	6.71	21194	21	7.68	0.86[c] (0.46-1.61)	0.87
	During Period After Stopping/Modifying Initial Treatment[d]							
	Patients Initiating Lithium			Patients Initiating Valproate				
Follow-up Time	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Odds Ratio (95% CI)	Rate Ratio
0-365 days	19494	38	7.58	19362	25	5.05	1.51[e] (0.91-2.50)	1.50

Table 2-3. (continued)

[a] p=0.32.

[b] The counts of patients "During Exposure to Initial Treatment" include all the propensity score-matched patients, since all patients accrued at least some follow-up time in that status. Counts of suicide deaths among these patients indicate suicide deaths occurring on a day in which the patient was classified as still receiving initial treatment. That is, these counts represent suicide deaths occurring during the period covered by a prescribed supply of medication (without any co-prescription of the other medication), or during the gap(s) permitted after the prescription had ended, up until the day that the first gap of 15 or more days had occurred.

[c] p=0.63.

[d] The count of patients "During Period After Stopping/Modifying Initial Treatment" indicates all the patients who reach that status by the end of the follow-up period, since all such patients accrued at least some follow-up time during which they were not still receiving their initially assigned treatment. That is, this is a count of patients modifying their initial treatment by switching to or augmenting with the other medication or discontinuing their initial treatment, either temporarily or permanently. Counts of suicide deaths among these patients among these patients indicate suicide deaths occurring on a day after the patient had exited "still receiving initial treatment" status, whether by discontinuing or modifying their initial treatment.

[e] p=0.11.

measure traditionally reported in many nonrandomized studies of lithium. Finally, increased suicide risks during the 0-365 day period were observed, although not statistically significant, among patients once they had stopped or modified lithium, compared to valproate, treatment (Odds Ratio (OR) = 1.51, 95% CI 0.91, 2.50, p = 0.11).

Additional secondary analyses (Table 2-4A and 2-4B) indicated an increased intent-to-treat risk of suicide of among patients initiating lithium over 0-180 days that was marginally statistically significant (Conditional Odds Ratio (cOR) = 1.56, 95% CI 0.92, 2.69, p = 0.08), in association with significantly elevated risk of suicide among patients after stopping or modifying lithium, compared to valproate, treatment during this period (Odds Ratio (OR) = 2.72, 95% CI 1.21, 6.11, p = 0.015). The differing risk of suicide between individuals stopping/modifying lithium treatment and stopping/modifying valproate treatment over 0-180 days was associated almost exclusively with those stopping, rather than modifying, treatment (Table 4, Footnote i: OR = 3.61, 95% CI 1.34, 9.73, p = 0.011).

Figure 2-1 presents the intent-to-treat survival curve for suicide for the lithium and valproate treatment groups. Because of nonproportional hazards (the crossing of the survival curves at approximately 90 days), the interpretation of the survival analysis (Appendix 1-4) is less straightforward than, but generally consistent with, the logistic regression results. Table 5 indicates that recent suicidal ideation, as reflected by diagnostic code, was significantly more prevalent among patients initiating lithium than among those initiating valproate (OR = 1.30, 95% CI 1.09, 1.54, p = 0.003) for the 19,411 patients for whom these data were available.

Table 2-4. Suicide Deaths and Rates over Time by Mood Stabilizer Treatment over Briefer Time Periods (0-90 and 0-180 days)

TABLE 2-4A. Intent-to-Treat Analyses

	All Patients Initiating Treatment (Intent-to-Treat Cohort)							
	Patients Initiating Lithium			Patients Initiating Valproate				
Follow-up Time	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Conditional Odds Ratio (95% CI)	Rate Ratio
0-90 days	21194	18	9.35	21194	19	9.87	0.95[a] (0.50-1.81)	0.95
0-180 days	21194	39	10.2	21194	25	6.54	1.56[b] (0.94-2.58)	1.56

TABLE 2-4B. Findings Stratified by Initial Treatment Status

	During Exposure to Initial Treatment							
	Patients Initiating Lithium			Patients Initiating Valproate				
Follow-up Time	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Conditional Odds Ratio (95% CI)	Rate Ratio
0-90 days	21194	15	10.0	21194	17	11.3	0.88[d] (0.44-1.77)	0.88
0-180 days	21194	17	7.99	21194	17	7.93	1.00[e] (0.51-1.96)	1.01

Table 2-4. (continued)

	During Period After Stopping/Modifying Initial Treatment[f]							
	Patients Initiating Lithium			Patients Initiating Valproate				
Follow-up Time	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Patients, No.	Suicide Deaths, No.	Rate (per 10^6 person-days)	Odds Ratio (95% CI)	Rate Ratio
0-90 days	11227	3	6.98	11185	2	4.75	1.49[g] (0.25-8.95)	1.47
0-180 days	16138	22	13.0	15958	8	4.78	2.72[h,i] (1.21-6.11)	2.72

[a] p=0.87.

[b] p=0.08.

[c] See Table 3, Footnote b.

[d] p= 0.72

[e] p>0.99.

[f] See Table 2-3, Footnote d.

[g] p=0.66.

[h] p=0.015

[i] Risks observed in patients stopping or modifying initial treatment were almost exclusively observed in patients stopping treatment (rather than modifying treatment or discontinuing and later resuming either treatment):

Patients Stopping Lithium: Suicides = 18; Suicide Rate (per 10^6 person-days) = 18.3.
Patients Stopping Valproate: Suicides = 5; Suicide Rate (per 10^6 person-days) = 5.09.

This yields an odds ratio of 3.61 (95% CI 1.34, 9.73) and a rate ratio of 3.60.

Figure 2-1. 365-day Survival Curve of Suicide by Treatment (Intent-to-Treat Analysis)

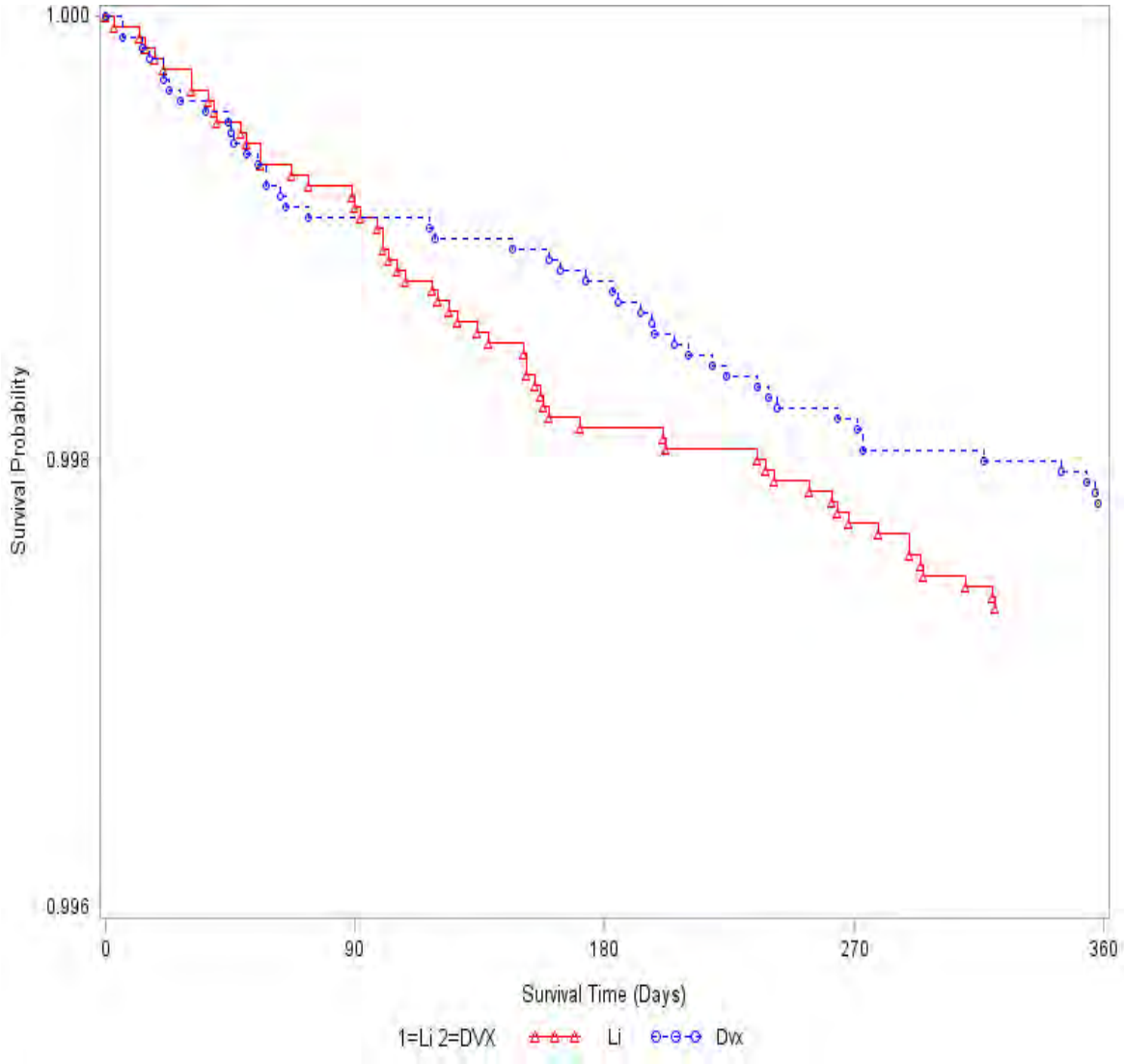


Table 2-5. Presence of V-code (62.84) denoting Suicidal Ideation in the 30 days prior to Lithium or Valproate Initiation (2005-2008)

Patient Characteristic	Patients Initiating Lithium	Patients Initiating Valproate	Odds Ratio (95% CI)	P value
Suicidal Ideation	305	237	1.30 (1.09-1.54)	0.003
No Suicidal Ideation	9478	9391		

Discussion

This manuscript reports the largest study, to our knowledge, examining lithium's association with suicide risk, and the first to use two design elements (propensity score matching and intent-to-treat analysis) intended to help nonrandomized studies better approximate findings from randomized trials. Lithium treatment, compared to valproate treatment, was not associated with reductions in suicide risk among VHA patients with mental health diagnoses over the first year of treatment, either in the primary intent-to-treat analyses or in secondary analyses of patients actively receiving their initial treatment.

This study's findings diverge from past meta-analyses.[12, 15] Several potential reasons suggest themselves. First, follow-up was only continued for one year (some studies have specifically reported that treatment with lithium for > 1 year was required to observe significant reductions in suicide risk).[67-69] A second reason may relate to characteristics of this sample (a Veteran sample with high rates of treatment discontinuation). High rates of treatment discontinuation would be expected to be especially influential in an intent-to-treat design, since effect estimates would substantially reflect risks observed during periods of nonexposure after discontinuation of treatment. Furthermore, in naturalistic studies high rates of discontinuation can complicate interpretation of risks even among patients apparently receiving active treatment, since it cannot be clearly ascertained whether or when patients prescribed medication consume it. Outcomes for patients who do not start a received prescription or

terminate it early may be ascribed to active treatment, but actually relate to the risks associated with nonexposure or treatment discontinuation. Alternatively, some prior studies reporting large associations between lithium treatment and reduced suicide risk may have been biased by inadequately controlled initial confounding or differences in selection occurring during treatment.[13, 14]

However, two other possibilities deserve consideration, since they are consistent with important elements of our data and would have implications for patients, providers, and healthcare systems. Both relate to the statistically significantly increased suicide risks observed among patients discontinuing lithium at 0-180 days. Risks among patients discontinuing treatment may result from several causes, including: 1) genuine new risks produced by medication discontinuation, and 2) intrinsic, non-treatment-related differences in suicide risk between the treatment groups (i.e., baseline confounding not resolved by the propensity score matching). A third potential cause, differing tendencies between the treatment groups for high-risk patients to discontinue treatment (through self- or provider-based selection), may also contribute to these risks but appears unlikely to explain the entirety of our findings (Appendix 1-5A). If risks observed after lithium discontinuation relate directly to discontinuing lithium, this finding would appear consistent with a substantial literature documenting pronounced suicide risks upon lithium discontinuation (e.g., suicide rates up to 14-fold greater than the rates observed during the prior lithium treatment).[13, 20, 21] However, most of these studies were uncontrolled. The only two prior studies directly comparing lithium and valproate

observed similar, not different, suicidal behavior risks among patients discontinuing either treatment.[22, 23]

The substantial increased risks observed among patients discontinuing lithium over 0-180 days could also reflect potential confounding remaining after propensity score matching[48, 49] (i.e., patients initiating lithium being at higher intrinsic suicide risk). If active lithium treatment was associated with reduced suicide risks, then those reduced risks might largely counterbalance confounding in the analyses of the intent-to-treat cohort and of the patients still receiving initial treatment, leaving such confounding to be revealed primarily after treatment discontinuation. It is unclear whether much confounding persists in our analysis, given the approximate initial balance observed in numerous measured factors and the further balance achieved after propensity score matching. Several lines of evidence suggests, however, that if any substantive confounding does exists, it likely biases against lithium. The effect of further increasing covariate balance through propensity score matching was to noticeably reduce initial effect estimates associating lithium with increased suicide risk. This suggests the initial imbalance in propensity score covariates, although generally small, biased associations towards observing higher risks for lithium. In addition, our data, while extensive, does not include information on several important risk factors (suicidal planning, means, recent stressors, and psychiatric symptoms, and, for some individuals, information on suicidal ideation). Our analysis of diagnostically-coded suicidal ideation found modest but significantly higher rates among patients initiating lithium, even after the propensity score matching.

A role for chance is also important to consider in interpreting our findings, since a substantial number of comparisons were examined and only three statistically significant associations were observed: among patients discontinuing lithium compared to valproate over 0-180 days (Table 4), among all patients initiating lithium compared to valproate from 91-180 days, and among patients discontinuing lithium compared to valproate over 91-180 days (Appendix 1-4). Nevertheless, while our primary findings over 0-365 days indicate no statistically significant differences between the treatments, results from even a study of this size do not preclude potential clinically meaningful differences existing between the treatments below the power of this study to detect. The significant risks in patients discontinuing initial lithium treatment over 0-180 days generally suggests some degree of nonequivalency between the treatments, with lithium being associated with distinct risks upon discontinuation compared to valproate and/or (if some or all of the risks associated with lithium discontinuation reflect confounding), with more positive benefit against suicide than suggested by our findings, especially during active treatment.

Several study limitations should be noted. Data limitations include gaps in prescription records for inpatients or patients receiving care outside the VHA, and potential errors in measurement of covariates. Analytic limitations included the absence of rematching/reweighting of patients during follow-up. This limitation precluded analysis of whether differences exist between the treatment groups during follow-up in either the initiation of, or persistence with, treatment with other psychiatric medications (e.g., antidepressants or antipsychotics). This study was a study of typical care, rather than being restricted to monotherapy, unlike some recent studies.[22, 23] Given that other

psychiatric medications may influence suicide risk,[70, 71] strategies such as marginal structural models which reweight patients during follow-up should be considered in the future (Appendix 1-5A). However, numerous classes of psychiatric medications prescribed at and before lithium or valproate initiation were very closely balanced between the two treatment groups, thus likely producing close similarity in concomitant medications, at least early during follow-up. The absence of rematching/reweighting also precluded a determination of whether patients experiencing ongoing or emergent suicidal ideation or behavior during follow-up were more likely to be discontinued from one of the two treatments.

In addition, serum medication levels would have provided information beyond simply prescription data about medication persistence, if these had been available. Study findings might have been influenced by the considerable diagnostic heterogeneity of this patient cohort, although each individual diagnosis was closely balanced in prevalence among lithium and valproate recipients and almost 90% of our sample had mood disorder diagnoses (Appendix 1-1). Our focus upon suicide mortality (comprehensively documented nationwide, even for patients who leave VHA care) improved outcome ascertainment compared to nonfatal suicidal behavior, but unfortunately limited statistical power. Generalizability to non-VHA patients, to patients with the excluded medical conditions (e.g., cancer, head injury, or seizures), and to cohorts with differing rates of treatment discontinuation or that are treated for longer than one year is uncertain.

Propensity score methods may also potentially inadvertently amplify any remaining confounding, primarily if variables are included that are substantially

associated with treatment (i.e., lithium or valproate initiation) but not outcome (i.e., suicide).[46, 59] How much of a bias is typically produced is controversial.[46, 72] Nevertheless, the differences between treatment groups in baseline suicidal ideation diagnoses suggests some unaddressed confounding may remain in this study, and our particular implementation of propensity score matching included extensive variables in some domains (e.g., medical diagnoses) in which only a subset of variables may have been strongly related to suicide risk. (However, Harris and Barraclough found that 90% of the medical diagnoses they reviewed were significantly associated with suicide risk).[64] To reduce the potential for amplification of remaining confounding (also termed "residual confounding"), we actively made judgments concerning inclusion/exclusion of the covariates that are of most concern (i.e., covariates with substantial associations with treatment). In addition, extremely few propensity score covariates (< 2%) exhibited an initial substantial association with treatment (≥ 0.1 initial standardized difference), suggesting that few variables were included that would contribute substantially to residual confounding amplification. Furthermore, a sensitivity analysis targeting this concern by removing a large number of variables (approximately half of the total) with the weakest apparent associations with suicide produced only modest effect estimate changes (Appendix 1-6). Most importantly, the direction of change in the 0-365 day effect estimates from the unmatched (OR = 1.45) to the matched sample (cOR = 1.22) strongly suggests that overall confounding was most likely reduced by the propensity score matching, not amplified (Appendix 1-5B). Therefore, any amplification of residual confounding appears to be sufficiently minor that the propensity score methodology still

produced an important reduction in overall confounding. When comparing this study to other studies, however, it is important to note that residual confounding amplification could have potentially enhanced, to at least a slight degree, an apparent bias in this study against observing a protective association for lithium treatment against suicide risk.

Nevertheless, our findings generally agree with the most recent randomized and nonrandomized studies. With one exception[73] recent nonrandomized studies of suicide or suicidal behavior risk have observed nonsignificant (and typically modest) differences between lithium and valproate, or lithium and anticonvulsants in general.[22, 23, 74-77] Recently, a small but methodologically-rigorous trial focused on suicidal behavior prevention[53] observed only nonsignificant differences in suicidal behavior between lithium and valproate. Results from this trial (involving 2.5 years of follow-up) and the BALANCE trial[27] (involving 2 years of follow-up) were combined in an updated randomized trial meta-analysis[25] which estimated that lithium treatment was associated with a nonsignificant reduction in nonfatal suicidal behavior compared to valproate (OR = 0.64, 95% CI 0.30, 1.36, p = 0.24). These findings appear broadly consistent with this study, in that all three studies observed nonsignificant intent-to-treat differences between lithium and valproate. The central estimate for lithium's effect size, however, did differ in direction between the three studies (nonsignificantly decreased suicidal behavior risk in the trials versus nonsignificantly increased suicide risk in this study). This difference might simply be due to chance, residual confounding (possibly augmented by residual confounding amplification), differences in follow-up time (one year versus 2-2.5 years), or differences in outcome (suicide versus suicidal behavior). However, this difference

could also reflect a "two-sided" nature to lithium's association with suicide risk. That is, some degree of decreased suicide/suicide behavior-related risk may be associated with active lithium treatment (and thus contributing more greatly to the trial intent-to-treat estimates, which had much higher treatment persistence rates), combined with some degree of increased risk associated with lithium discontinuation (which would thus contribute more greatly to the intent-to-treat estimates in our study). This possible "two-sided" association between lithium and suicide risk would be also consistent with the significant differences in baseline suicidal ideation diagnoses between the treatment groups observed in this study (suggesting some residual confounding and thus a greater benefit to active lithium treatment than indicated) and the timing of the emergence of significant risks after lithium discontinuation (suggesting risks associated with discontinuation itself) (Appendix 1-5E).

Clearly, our findings illustrate a need for further research (Appendix 1-7). Until such research resolves whether the increased risks observed in patients discontinuing lithium relate specifically to lithium discontinuation, prudence suggests patient and provider education about the possible risks of lithium discontinuation and close monitoring of patients discontinuing lithium (and, potentially valproate[22, 23] when feasible. Such monitoring is already recommended to limit mood episode recurrences.[78] In addition, healthcare systems, providers, and patients should strive to maximize persistence with lithium treatment once initiated. When discontinuation does occur, there may be value to facilitating a gradual discontinuation of lithium by patients when appropriate.[79, 80]

Conclusions

In summary, this study did not observe significant benefits for lithium in preventing suicide compared to valproate among Veterans Health Administration patients over the first year of treatment. This study is notable, however, for high rates of discontinuation of both lithium and valproate, and for the finding of increased suicide risk among patients discontinuing lithium over 0-180 days. If such increased risk largely reflects confounding still persisting in the analysis, such confounding could conceal a clinically meaningful suicide preventative effect for lithium. Alternatively, some or all of the risk among patients discontinuing lithium could represent genuinely greater risks of suicide related to lithium, compared to valproate, discontinuation. Until further research more fully clarifies the relationships between lithium treatment, discontinuation, and suicide, patients initiating lithium should be educated concerning the possible risks associated with lithium discontinuation and the need to maximize persistence with lithium treatment, and receive close monitoring after discontinuation if feasible. Further research incorporating intent-to-treat approaches is clearly needed, given the possible beneficial or hazardous effect sizes still compatible with this study's results, the pressing need for interventions against suicide, and the broad potential use of lithium.

Chapter III

Nonsuicide Mortality Associated with Lithium and Valproate Treatment of US Veterans Health Administration Patients with Mental Disorders

Abstract

Background/Aims: To assess associations between lithium, valproate, and nonsuicide mortality.

Methods: Intention-to-treat, high-dimensional propensity score-matched incident-user cohort study of Veteran's Health Administration (VHA) patients with mood or psychotic disorders newly initiating lithium or valproate from 1999-2008 using Cox regression (n=21, 288/treatment).

Results: Matching produced treatment groups closely similar in every one of an extensive set of measured covariates (all standardized differences < 0.019). Significant differences between lithium and valproate were not observed for our primary, intent-to-treat analysis over 0-365 days (HR = 0.92, 95% CI 0.82, 1.04). Lithium initiation was associated with significantly reduced nonsuicide mortality in the intent-to-treat cohort for the secondary endpoint of 0-90 days (HR = 0.67, 95% CI 0.51, 0.87). In other secondary analyses, a sizeable reduction in mortality was observed during active treatment with lithium across all time periods studied (e.g., 365-day HR = 0.62, 95% CI 0.45, 0.84), but

significantly increased risks were observed among patients discontinuing lithium by 180 days (HR = 1.54, 95% CI 1.01, 2.37).

Conclusions: Patients initiating lithium had did not have substantially lower nonsuicide mortality than patients initiating valproate for our primary endpoint of 0-365 days. In secondary analyses, lower hazards of nonsuicide mortality were observed for patients initiating lithium compared to valproate over 0-90 days (intent-to-treat analysis). Consistently lower nonsuicide mortality was among patients maintaining treatment, but elevated risk among patients discontinuing treatment by 180 days. While residual confounding or selection effects cannot be excluded, this study suggests potential benefits to enhancing lithium treatment persistence and monitoring of patients discontinuing lithium, and identifies important needs for further research.

Introduction

While a substantial literature exists concerning the potential association of lithium with reduced suicide mortality, very few studies have examined lithium's potential influence on nonsuicide mortality. Many organ systems are exposed to lithium,[81, 82] and lithium produces diverse physiological effects. Some of these effects are potentially beneficial (e.g., leukocytosis,[32] reduced heart rate,[33] and neurogenesis[34]), while others are potentially hazardous (e.g., renal[35, 36] and thyroid insufficiency,[37] QTc prolongation,[33] or arrhythmias[38-40]). A limited number of studies, primarily[83-90] but not exclusively[12] nonrandomized, have examined associations between lithium treatment and nonsuicide mortality. These studies are consistent with the possibility that lithium might reduce mortality risk in psychiatric patients. Determining the effects upon nonsuicide mortality of lithium or other psychiatric treatments is clearly important, especially since patients with serious mental illness are at particular risk for premature mortality.[91-94]

We conducted a nationwide cohort study of the United States (US) Veterans Health Administration's (VHA) detailed clinical databases, employing two methods intended to increase the likelihood that observational studies will yield results similar to randomized trials: high-dimensional propensity score matching and intent-to-treat estimates. High-dimensional propensity score matching permit inclusion of particularly detailed information concerning potential confounding while facilitating the assessment of the balance in these potential confounders that is achieved between treatment groups. Intent-to-treat estimates enhance interpretation of results by allowing assessment of

whether benefits during active treatment are negated by risks upon discontinuation. Employing these approaches, we investigated whether initiation of lithium was associated with reduced nonsuicide mortality compared to initiation of valproate, a treatment which has largely replaced lithium in many countries.[95-98]

Methods

Data Sources

Demographic characteristics, inpatient and outpatient mental and non-mental health treatment records, and outpatient pharmacy prescription data was obtained from the VHA National Psychosis and Depression Registries.[54] (These registries are linked, de-identified healthcare databases of all VHA patients nationwide since 1997 with at least one psychotic or depressive disorder diagnosis). This study was approved by the Institutional Review Boards of the Bedford and Ann Arbor Veterans Affairs Medical Centers.

Study Cohort

Incident users[50] (≥ 6 months of no lithium or valproate use but with recent VHA utilization) receiving at least one outpatient prescription for lithium or valproate from April 1999 to December 2008 were identified (Appendix 2-1). A broad cohort of patients with mood or psychotic diagnoses in the 30 days prior to medication initiation was examined since the limited prior literature concerning lithium and mortality is not restricted to bipolar disorder (Appendix 2-2).[12, 84, 86, 90]

Patients with possible non-psychiatric indications for valproate or lithium (epilepsy, migraine or cluster headache, or neuropathy diagnoses in the past 30 days; dementia medication use in the past 180 days; cancer, dementia, skull fracture diagnosis, traumatic brain injury diagnosis or treatment, home care, or hospice care in the past year; or any nursing home residence or inpatient rehabilitation in the past two years) were excluded. Patients were also excluded if they initiated lithium or valproate on an "as needed" basis, both medications simultaneously, or resided outside the United States.

Exposure Determination

Receipt of lithium or valproate was defined by a fill of an outpatient prescription for these medications. For the intent-to-treat analysis, all patients filling an initial outpatient prescription were followed until end of follow-up (i.e., 90, 180, or 365 days) or death.

Secondary analyses stratified follow-up time by whether patients were still receiving initial treatment. Patients were considered "as-initially-treated" until a ≥15 day gap occurred between outpatient prescriptions (adjusting for early refills), or upon initiation of the other mood stabilizer (i.e., lithium or valproate). Patients were considered "former users" for the period of time after initial treatment discontinuation until follow-up time was censored upon subsequent treatment resumption or switching (to the other treatment), death, or the end of follow-up. Given that this "former user" follow-up period is free from exposure to either medication studied, former users have been advanced as a possible index of the potential biasing effects of residual baseline confounding and/or selection occurring during treatment.[48, 49] However, risks among

former users can be more fully conceptualized as representing the combined effects of residual confounding and selection along with any persistence of effects from active treatment, and any risks produced upon treatment discontinuation (Appendix 2-3). These "discontinuation-associated risks" would include effects such as "rebound" mania or depression.

Outcome

Date and cause of death was obtained from National Death Index files for 1999-2009.[56] This study was limited to nonsuicide mortality, with follow-up time for patients dying of suicide censored at suicide death.

Propensity Score Modeling

948 covariates derived from VHA databases were included in an initial propensity score model generally following the "high-dimensional" propensity score approach[41, 58] (Appendix 2-4). These covariates included potential risk factors for both non-suicide and suicide mortality[99] (including demographic characteristics, diagnoses, general VHA mental and non-mental health services utilization,[41] hospitalizations, clinic use, occurrence of diagnostic testing, current and recent prescriptions, recent injuries and diagnosed suicide attempts, and state-level and VHA-hospital subsystem mortality risk[100]) (Appendices 2-5 and 2-6), often with multiple indicator variables to allow for nonlinear covariate-mortality relationships. An "outcome-focused" propensity score was then derived limiting covariates to the 523 covariates with substantial associations with outcome[101] (i.e., +/- 20% change in nonsuicide mortality).[102] Further details of how this

covariate restriction was implemented are provided in Appendix 2-6. This outcome-focused propensity score, intended to limit unintended amplification of confounding which remained uncontrolled, [46, 59, 101] provided the basis for the results reported here. Results from analyses using the initial propensity score are provided in Appendix 2-7. The results from the two models are generally similar but differ in a few important details, such as the time periods for which significant associations are detected.

Statistical Methods

The propensity score was calculated using logistic regression. Patients initiating lithium and valproate were 1:1-matched (Appendix 2-8) using calipers of 0.2 standard deviations of the propensity score logit,[60,61] resulting in 99.3% matching of lithium-initiated patients. Balance in covariates between treatment groups was assessed using standardized differences (equivalent to Cohen's *d* effect sizes, with a difference of > 0.10 indicating significant imbalance).[42]

Statistical significance was determined using techniques that reflected matching (stratified Cox regression with sandwich variance estimators) for the primary intent-to-treat analyses and the secondary as-treated analyses. Ordinary Cox regression was used for the secondary former user analysis (since matching was not preserved for this analysis). All analyses except standardized differences performed using SAS, version 9.3. Standardized differences were calculated using Microsoft Excel 2010.

Results

The incident user cohort of 93,162 patients initiating lithium or valproate was generally balanced (standardized difference < 0.10)[42] between the treatment groups in virtually all non-mental health and mental health covariates even prior to matching. After matching, substantial additional balance was achieved between treatment groups (n=21,288 patients per group). As an end result, high-dimensional propensity score (hdPS) matching achieved a very close balance (standardized differences <0.019) for each one of the 523 covariates included in the outcome-focused analysis. Table 3-1 indicates the standardized differences in the matched cohort for two categories of covariates: 1) those few variables with a substantial imbalance (≥ 0.10 standardized difference) between treatment groups initially, and 2) a number of additional covariates with well-established or highly plausible relationships with nonsuicide mortality. These additional covariates include age, sex, disability status, recent number and types of hospitalizations, and particular diagnoses, medications, and attendance at certain outpatient clinics. Comparison of the unmatched and hdPS-matched effect estimates indicate that hdPS-matching reduced effect sizes in a direction consistent with reducing baseline confounding biasing against valproate (Appendix 2-9).

Impersistence with treatment was very common even within 180 days, but rates of treatment impersistence were highly similar between the treatment groups: 76.4% of patients initiating lithium and 75.4% initiating valproate did not persist with initial treatment for 180 days (Appendix 2-10).

Table 3-1. Characteristics of Patients Initiating Lithium and Valproate (Propensity-score Matched Sample)

Patient Characteristic	Lithium (n=21288) N	Lithium (n=21288) (%)	Valproate (n=21288) N	Valproate (n=21288) (%)	Standardized Difference
INITIALLY SUBSTANTIALLY IMBALANCED COVARIATES (Initial Standardized Difference ≥ 0.10)					
Sex (Female)[a]	2932	13.8	2952	13.9	-0.003
Bipolar I Disorder, past 30d	9630	45.2	9719	45.7	-0.008
Other Psychosis, past30d	251	1.2	261	1.2	-0.004
Post-Traumatic Stress Disorder (PTSD), past year	4858	22.8	4849	22.8	0.001
Other Mood Stabilizer(s), current	2963	13.8	2894	13.6	0.006
Prior Mood Stabilizer Treatment	7573	35.6	7473	35.1	0.010
Mild Liver Disease, past year	1495	8.4	1708	8.0	0.015
ACE Inhibitor, current	2772	13.0	2765	12.9	0.001
SELECT ADDITIONAL VARIABLES WITH LESSER INITIAL IMBALANCES					
Age, 65-79[b]	1358	6.4	1359	6.4	0.000
Age, 80+[b]	164	0.8	154	0.8	0.005
Married	7455	35.0	7370	34.6	0.008
Disability (51-100%)	5473	25.7	5481	25.7	-0.001
Charlson Comorbidity Index, past year, 1+	7601	35.7	7468	35.1	0.013
Myocardial Infarction, past year	235	1.1	243	1.1	-0.004
Diabetes (Uncontrolled), past year	2722	12.8	2691	12.6	0.004
Arrhythmia, past year	860	4.0	856	4.0	0.001
Chronic Obstructive Pulmonary Disease, past year	2968	13.9	2934	13.8	0.005
Total Nonpsychiatric Medications, current, 5+	5670	26.6	5589	26.3	0.009
Beta-Blockers, current	2757	13.0	2713	12.7	0.006
Opioid Pain Medication, current	2409	11.3	2374	11.2	0.005
Antiplatelet Agent, current	221	1.0	218	1.1	0.001
Warfarin, current	200	0.9	187	0.9	0.006
Nonpsychiatric Hospitalizations, past year, 1+	1842	8.7	1781	8.4	0.010
Latest Discharge from Medical ICU	330	1.6	323	1.5	0.003
Latest Discharge from Neurology	54	0.3	54	0.3	0.000

Table 3-1. (continued)

Nonpsychiatric Discharge, Against Medical Advice (AMA), past year	207	1.0	203	1.0	0.002
General Surgery, past 180d,[c] 1+	705	3.3	697	3.3	0.002
Nonpsychiatric Visits, last 7d,[c] 1+	4479	21.0	4426	20.8	0.006
Specialty Visits, past 180d,[c] 1+	3720	17.5	3619	17.0	0.013
Cardiology clinic, past 180d,[c] 1+	594	2.8	599	2.8	-0.001
Pain clinic, past 180d,[c] 1+	447	2.1	467	2.2	-0.006
Chaplain Service, past 180d,[c] 1+	547	2.6	540	2.5	0.002
Nuclear Medicine, past year, 1+	837	3.9	812	3.8	0.006
Alcohol Dependence	4440	20.9	4449	20.9	-0.001
Heroin/Opiate Dependence	824	3.9	810	3.8	0.003

[a] The sex ratio of our sample differs substantially from that of bipolar disorder in the general population because our sample is a Veteran sample, primarily made up of male individuals. It is particularly important to note that the hdPS-matching procedure did not substantially change the prevalence of male or female sex in the final study cohort compared to the initial sample, or the prevalence of any of the other included covariates. Rather, the matching procedure led to the selection of a set of patients initiating valproate who were very similar to those initiating lithium. For instance, in the original study cohort prior to matching 13.9% of patients initiating lithium and 9.4% of patients initiating valproate were female. In the final cohort, 13.7% of patients initiating lithium and 13.7% of patients initiating valproate were female.

[b] Age presented in this format (65-79 years old and 80+ years old) because these were the age groups with highest mortality. Age was actually modeled using 11 indicator variables reflecting age groups from < 35, 35--39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, and 80+ years old.

[c] "d"= days; e.g. 30d = 30 days.

Overall survival was greater among patients initiating lithium, with 274 deaths over 365 days observed among the lithium intent-to-treat cohort, compared to 296 deaths among the valproate intent-to-treat cohort. Greater differences (71 versus 101 deaths, respectively) were observed between the as-treated cohorts. Survival curves for the intent-to-treat and as-treated analyses are provided in Figures 3-1A and 3-1B.

Table 3-2 provides the primary, intent-to-treat analysis results, indicating that lithium was associated with substantially reduced mortality risks over 0-90 days (hazard ratio (HR) = 0.67, 95% CI 0.51, 0.87), the period of greatest medication persistence, but not 0-180 days (HR= 0.97, 95% CI 0.82, 1.15) or 0-365 days (HR = 0.92, 95% CI 0.82, 1.04). Secondary analyses by treatment status (Table 3-3) reveal large and significant associations with nonsuicide mortality during active lithium treatment compared to valproate treatment over all time periods. Hazard ratios were consistently and considerably lower during the period of likely active use of lithium compared to likely active use of valproate (as-initially- treated hazard ratios ranging from HR = 0.59, 95% CI 0.42, 0.84, to HR = 0.62, 95% CI 0.45, 0.84). However, significantly increased nonsuicide mortality was also observed among lithium former users (n = 54) than valproate former users (n=35) over 0-180 days (HR = 1.54, 95% CI 1.01, 2.37), although not for other time periods.

Table 3-4 indicates no significant intent-to-treat associations existed between treatment and specific categories of causes of death at 365 days. The mortality categories with associations closest to statistical significance, however, were cardiovascular disease (CVD) and deaths from all other causes. Upon further examination, these categories

Figure 3-1A. Survival curve of lithium and valproate treatment, intent-to-treat cohort

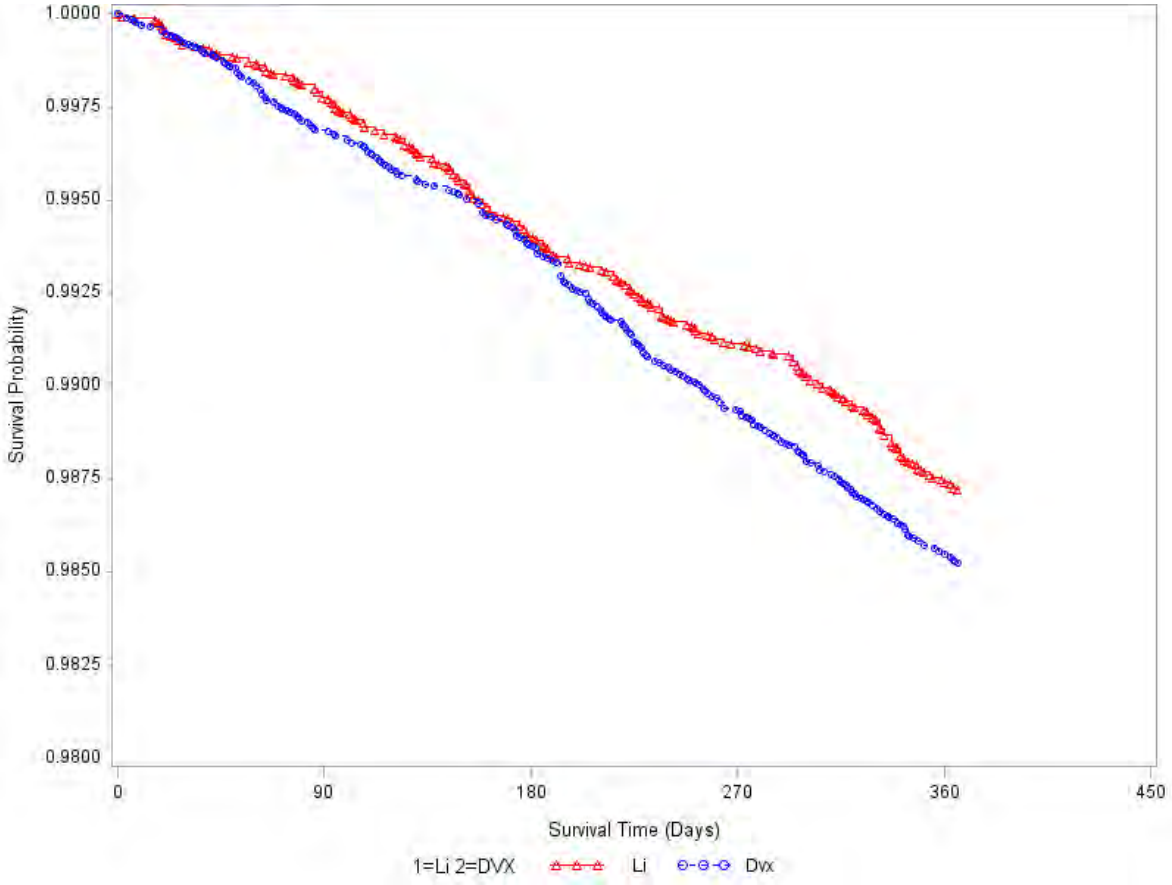


Figure 3-1B. Survival curve of lithium and valproate treatment, as-initially-treated patients

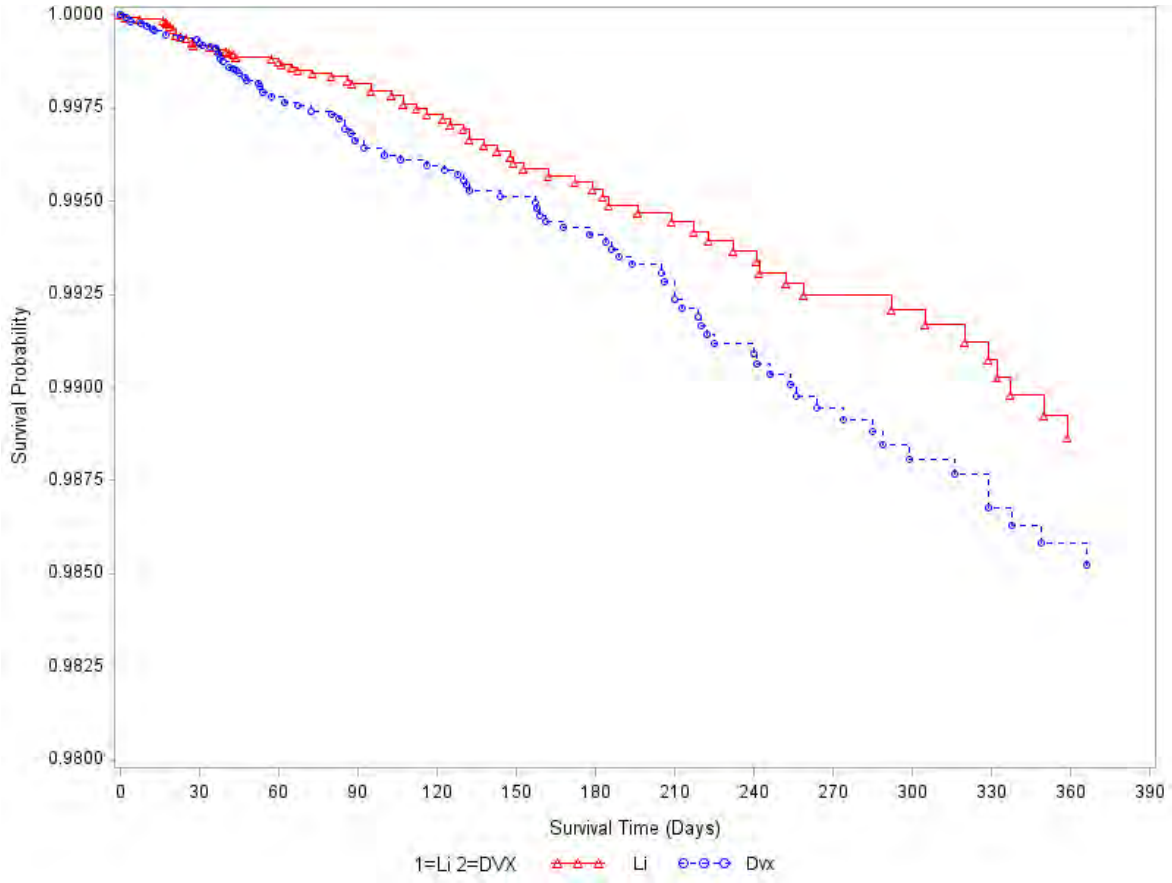


Table 3-2. Risk of Nonsuicide Mortality (Intent-to-Treat Cohort)

Time Period	Hazard Ratio (Lithium/Valproate)
0-90 Days	0.67[a] (0.51-0.87)
0-180 Days	0.97[b] (0.82-1.15)
0-365 Days	0.92[c] (0.82-1.04)

[a] Based on 48 deaths / 1934388 person-days for lithium and 72 deaths / 1933337person-days for valproate; p = 0.003
[b] Based on 128 deaths / 3839959 person-days for lithium and 132 deaths / 3838384person-days for valproate; p = 0.73
[c] Based on 274 deaths / 7733701 person-days for lithium and 296 deaths / 7729420 person-days for valproate; p = 0.17

Table 3-3. Risk of Nonsuicide Mortality (Stratified by Exposure Status)

Time Period	Hazard Ratio (Lithium/Valproate)	
	During Initial Exposure (As-Initially-Treated)	During Subsequent Nonexposure (Former Users)
0-90 Days	0.59[a] (0.42, 0.84)	0.88 (0.45-1.74)
0-180 Days	0.59[b] (0.42, 0.82)	1.54[c] (1.01-2.37)
0-365 Days	0.62[d] (0.45, 0.84)	1.02 (0.79-1.32)

[a] p = 0.004
[b] p = 0.002
[c] p = 0.045
[d] p = 0.002

Table 3-4. Risk of Nonsuicide Mortality by Cause (Intent-to-Treat Cohort)

Cause of Death	Intent-to-Treat Hazard Ratio (Lithium/Valproate)
Cardiovascular (CVD) (n=171 deaths)	0.86 (0.70-1.06)
Injury (n=105 deaths)	0.94 (0.72-1.24)
Cancer (n=54 deaths)	1.25 (0.85-1.83)
Stroke (n=21 deaths)	1.20 (0.66-2.18)
All Other Causes (n=231 deaths)	0.87 (0.71-1.05)

were also the only categories to have significant (All other Causes, HR = 0.50, 95% CI 0.28-0.91) or borderline significant (CVD, HR = 0.60, 95% CI 0.36-1.01) associations among as-initially-treated individuals.

Discussion

In a nationwide cohort study of 42,576 VHA psychiatric patients initiating lithium and valproate, no significant difference in mortality risk over 0-365 days was observed in our primary, intent-to-treat analysis. However, interpretation of this finding is complicated by the fact that high rates of treatment impersistence were observed among individuals initiating either lithium or valproate. By 365 days, 92% of patients had discontinued their initial treatment, and approximately 75% had discontinued their treatment by 180 days. Such high rates of treatment impersistence make detecting intent-to-treat differences between treatments more unlikely. In contrast, significant intent-to-treat associations of lithium initiation with lower mortality risks were observed over 0-90 days. This is the period during which treatment persistence was highest, and the likelihood of detecting intent-to-treat differences expected to be the greatest. However, this period is also the time during which any initial confounding is expected to be the most pronounced.[103]

Although our high-dimensional propensity score successfully achieved close balance on a large number and variety of important potential confounders, some degree of remaining confounding remains plausible (Appendix 2-11). Risk estimates indicating greater mortality risk among patients initiating valproate in the unmatched sample

(Appendix 2-9) become less pronounced after hdPS-matching, suggesting that overall confounding initially biases against valproate. Intent-to-treat estimates are important not only because they are the preferred effect estimates in randomized trials, but also because in both randomized and nonrandomized studies, they remove the possibility of confounding arising after treatment initiation. As a result, the firmest conclusion that can be reached from our intent-to-treat analyses is that initiating lithium may be associated with reduced mortality, given the 0-90 day secondary intent-to-treat findings, if baseline residual confounding is minimal or modest. However, this intent-to-treat association is not maintained over longer follow-up, either due to the high rates of treatment discontinuation, the 0-90 day findings being a chance result, or, possibly, due to the occurrence of greater mortality risks associated with lithium than valproate discontinuation.

Our secondary ("as-initially-treated" and "former user") findings are consistent with both an association between active lithium treatment and reduced mortality risk, and an association between lithium discontinuation and increased mortality risk, compared to valproate. Among patients who persisted with initial treatment, strong associations were observed between lithium, compared to valproate, treatment and reduced mortality across all time periods. Even if confounding biasing against valproate exists, the intent-to-treat estimate both over 0-90 days (central estimate HR = 0.67) and the "as-initially-treated" effect estimate (central estimate HR = 0.59 to 0.62) are of such size to raise at least some question whether residual confounding could plausibly explain this entire association. Equally important to recognize, however, is that if any confounding does exist and it

generally bias against valproate, this suggests the increased mortality risk associated with patients discontinuing lithium from 0-180 days may be even greater than indicated.

Although not focused upon here, another potential contributor to the "as-initially-treated" and "former user" findings is differences between the treatment groups in which patients are selected during follow-up to discontinue versus remain on their treatment. A substantial contribution from differential selection during follow-up is consistent with some, but not all, of the study findings (Appendix 2-13). Thus, a potential integration of our findings is suggested in which some degree of mortality benefits during active lithium treatment are counterbalanced by some degree of mortality risks during discontinuation of lithium treatment, compared to valproate treatment. Assessing the precise likelihood of this scenario, however, or whether lithium initiation is associated with net benefit or harm over the first 365 days of treatment, is difficult.

In spite of these uncertainties, one clear clinical recommendation can be made: once initiated, persistence with lithium treatment should be monitored and if clinically indicated, maintained. Regardless of whether the predominant association of lithium treatment (compared to valproate) with mortality is one of lower mortality risks during active treatment or higher mortality risks during lithium discontinuation, in either case maximizing persistence with lithium treatment would be of clear benefit (Appendix 2-14).

Our results are generally consistent with a limited prior literature. A clinical trial meta-analysis of both placebo and comparator-controlled trials reported significant reductions in overall mortality with lithium treatment (HR = 0.42, 95% CI 0.27, 0.81).

However, the randomized trial results from comparator-controlled trials are informed by just 4 deaths among lithium recipients and 12 among comparator recipients (Reference 20, Figures 2 and 4). Nonrandomized studies of lithium's effects on nonsuicide mortality are few but generally indicate reduced risks with active lithium treatment, although they typically lack active comparators, intent-to-treat designs, or detailed controls for confounding.[83-90] Our study is, to our knowledge, the first nonrandomized active-comparator study to include an examination of nonsuicide mortality risks associated with lithium discontinuation. Nevertheless, the results here are consistent in a general sense with limited prior uncontrolled studies, which observed that lithium discontinuation is a high-risk period for overall mortality.[104, 105] This study's findings potentially could also be broadly consistent with prior randomized[106] and nonrandomized[79, 107] literature indicating that lithium discontinuation substantially increases risks of mood episodes. Finally, our conclusions concerning the importance of persistence with lithium treatment are generally consistent with multiple prior studies reporting substantial lithium treatment impersistence.[108-114]

Study limitations include our lack of inpatient prescription information, lack of serum medication levels as an alternative method to assess persistence with treatment, and the inherent inability to completely model potentially important covariates such as hospitalizations (Appendix 2-15). A few variables found important in past mortality studies (income and race) which are sometimes poorly measured in VA data were not included in the outcome-focused propensity score. While available medical information was extensively represented, this information was only present for treatment received at

the VHA. Although we employed multiple methods to attempt to balance the treatment groups in VHA medical utilization (including indicators such as the presence and number of recent nonpsychiatric medications, overall visits, and specialist visits a patient received), this lack of outside healthcare data may be particularly important for patients receiving emergency care (which is more likely to occur at the nearest available hospital) or for older patients with Medicare. We also did not rebalance our treatment groups during follow-up for time-varying factors such as medications received through methods such as marginal structural models, although the treatment groups were closely balanced on a very extensive set of psychiatric and nonpsychiatric medications present at treatment initiation. Given that a very large majority of patients had stopped or modified their initial treatment by 365 days, we did not examine patient outcomes occurring over >365 days from treatment initiation. Nevertheless, to the degree that either lithium or valproate is associated with health risks or benefits that accrue over > 1 year of treatment, the impact of these risks and benefits upon mortality will not be reflected in this study.

Patients with several major mental health diagnoses were included to achieve sufficient power (e.g., depression, bipolar, and psychotic diagnoses). Although the psychiatric diagnoses were each balanced closely between treatments by hdPS-matching, this may have introduced some heterogeneity in the associations between treatments and mortality. Suicide deaths, which some studies have reported as strongly influenced by lithium treatment[16, 29] and/or its discontinuation,[20] may have been miscoded to some extent as accidents/injuries, resulting in an outcome not completely specific for nonsuicide mortality. Studies of overall mortality have also been criticized in general for

their lack of specificity.[115] However, an overall nonsuicide mortality focus for this study appears appropriate, given that lithium and valproate affect so many organ systems that *a priori* cause-specific hypotheses are difficult.

Our study examines a U.S. Veteran sample, and as such its generalizability to non-Veteran samples may be uncertain. For instance, 86% of our sample is male. Perhaps less obviously, the intent-to-treat estimates that are essential for developing a full view of the possible risks and benefits of treatment produce important additional limitations to generalizability. For the intent-to-treat estimates to likely generalize to other patient samples, that patient sample would need to exhibit a similar rate of treatment persistence. The treatment impersistence rates observed here, however, appears to be quite consistent with those observed in the only other incident cohort from a broad sample of United States patients that we were able to identify. Johnson and McFarland reported a median time to discontinuation of the first episode of treatment with lithium of only 72 days in a U.S. Health Maintenance Organization sample.[113] Finally, although work in propensity score methodology has been steadily advancing, it has not been definitively determined whether outcome-focused propensity scores should be favored over larger propensity scores in all circumstances. Reassuringly, the results given here for the outcome-focused model and in Appendix 2-7 for the initial model are generally consistent in many aspects. These aspects include significant intent-to-treat difference between lithium and valproate at 90 days and substantial effect sizes for as-treated and former users that are almost uniformly consistent with the outcome-focused model in the direction of effects, although not always identical in significance.

In our judgment, this study clearly establishes high-priority clinical and research agendas. The data from this study clearly suggests a need for clinical systems and providers to encourage patients to continue with their lithium treatment. A limited literature exists concerning psychosocial interventions that might help accomplish this task.[116] In addition, two trials that have included group psychoeducation with an emphasis on educating patients on the importance of medication treatment and/or adherence for successful management of bipolar disorder have shown superior outcomes to standard care.[117, 118] Our data also supports monitoring patients closely upon discontinuation when feasible, a practice already recommended in some guidelines to limit mood episode recurrence.[78] Finally, some approaches such as gradual discontinuation[79] have been proposed to limit the adverse psychiatric effects of lithium discontinuation.

From a research perspective, this study establishes a need for further research to elucidate the balance of risks and harms associated with lithium initiation. These studies might include the use of instrumental variables to reduce confounding from imperfectly measured or unmeasured factors (if valid instruments can be identified), and/or marginal structural models to reduce the impact of differential selection during treatment on the secondary (i.e., "as-initially-treated" and "former users") analyses. Marginal structural models would also allow the impacts of additional treatments commenced during follow-up to be evaluated.

Whether any mortality differences between lithium and valproate are primarily due to their direct psychiatric effects (stabilizing mood), indirect effects on physical

health (e.g., mood stability possibly leading to better adherence to medical treatment), or direct effects on physical health (both medications affect many organ systems) remains to be elucidated. The plausibility of differences in psychiatric effectiveness contributing to nonsuicide may be supported by some recent randomized[27] and nonrandomized[119] studies that have reported greater efficacy or effectiveness for lithium that valproate for bipolar disorder. In the BALANCE trial, the valproate treatment arm underperformed both the lithium-valproate combination and the lithium alone treatment arms.[27] In Denmark, lithium was found to be associated with fewer subsequent psychiatric hospitalizations than valproate.[119] A few other interventions targeting mental health have been associated with changes in overall mortality,[120] but others have not.[121]

Conclusions

This cohort study of U.S. Veterans Health Administration patients observed significantly reduced nonsuicide mortality among all patients initiated on lithium compared to valproate over 0-90 days but not beyond this period. Furthermore, significant associations were observed in opposite directions in secondary analyses: reduced mortality associated with individuals receiving lithium treatment, and increased mortality associated with individuals discontinuing lithium treatment (over 0-180 days), relative to valproate. This pattern suggests a dual aspect to the associations of lithium and valproate treatment with mortality: associations in a beneficial direction associated with active lithium treatment that potentially exceed or are exceeded by counterbalancing by mortality associations in a harmful direction associated with lithium discontinuation.

Intrinsic uncertainties common to nonrandomized studies (e.g., confounding), despite our efforts to minimize them, preclude a definitive judgment of whether lithium initiation was associated with net mortality benefit or net harm compared to valproate initiation. One clear and important clinical conclusion nevertheless emerges: once lithium treatment has been initiated, patients and providers should strive to maximize persistence with lithium treatment when feasible and clinically indicated. Such a conclusion results regardless of whether lithium is associated with benefits during active treatment or harms after discontinuation. In addition, given the problem of premature mortality in patients with serious mental illness,[91-94] the potential mortality differences between lithium and valproate should immediately receive greater research attention.

CHAPTER IV

Suicide Risk in Veteran Health Administration Patients with Bipolar Disorder Initiating Lithium or Valproate

Abstract

Objective: Past literature has reported sizable associations between lithium treatment and reduced suicide risk in patients with bipolar disorder, but these studies have often lacked extensive controls for confounding, intent-to-treat designs, active comparators, and/or fixed and equal follow-up time between treatments.

Method: Intention-to-treat, high-dimensional propensity score-matched incident-user cohort study of Veteran's Health Administration (VHA) patients with bipolar disorder newly initiating lithium or valproate from 1999-2008 using Cox regression (n=11, 298/treatment).

Results: Matching produced treatment groups closely similar in every one of an extensive set of measured covariates (all standardized differences <0.024). Initiation of lithium treatment was associated with significantly higher risk of suicide than valproate initiation over the first year of treatment (0-365 day HR = 1.50, 95% CI 1.05, 2.15, p = 0.008) in our primary analysis. However, secondary analyses indicated that no increase in suicide risk was associated with patients still receiving their initial lithium treatment (365 day HR = 1.0, 95% CI 0.52, 1.92). In contrast, an increased risk of suicide was

observed subsequent to the discontinuation of lithium, compared to valproate, treatment although this association was only significant at 0-180 days (HR = 6.10, 95% CI 1.37, 27.3, p = 0.018).

Conclusions: Over the first 12 months after treatment initiation in this large, extensively-matched cohort study of VHA patients, lithium initiation was associated with higher suicide risks than valproate initiation. This association was largely observed among patients discontinuing lithium, compared to discontinuing valproate, treatment. High rates of treatment impersistence also meant that the intent-to-treat effect estimate predominantly reflects associations that were observed after treatment discontinuation. Several lines of evidence suggest some residual confounding may exist biasing towards observing higher suicide risks among patients initiating lithium. Nevertheless, our findings indicate a need for caution concerning employing lithium as a suicide preventative among bipolar patients, a need to educate patients and providers about potential risks associated with lithium discontinuation, and a need to develop approaches to maximize persistence with lithium treatment and minimize risks upon discontinuation.

Introduction

Reducing suicide is both a national priority and a particular need for Veterans.[1] The mood stabilizer lithium has long been reported to be associated with uniquely large reductions in risks of suicide and suicidal behavior.[15, 16, 29] Many studies, however, have been nonrandomized and lacked intent-to-treat designs, extensive controls for confounding, active comparators and/or restrictions to incident-users.[14] Evidence from randomized trials is extremely limited (e.g., a meta-analysis of 32 trials included just 2 suicides among patients assigned lithium), but consistent with large, significant reductions in suicide risk (OR = 0.26).[12]

In contrast, recent nonrandomized studies[22, 23, 74-76] and a small randomized trial of suicidal behavior[53] report smaller, nonsignificant reductions in suicide or nonfatal suicidal behavior associated with lithium in comparison to another commonly-used mood stabilizer, valproate.

While nonrandomized studies can provide the large sample sizes desirable in studies of rare events such as suicide, concerns exist that earlier studies may have been confounded by prescription of lithium to patients preferentially at lower suicide risk.[13, 14] Over the past 10-15 years, however, prescriber behavior may have changed given well-publicized meta-analyses,[10, 20] treatment guidelines,[11] and high-profile studies[16] that have reported lithium treatment to be associated with distinct reductions in suicide risk.

Focusing on this more recent time period, we conducted an important follow-up study to our investigation examining lithium and valproate treatment and suicide risks among Veterans Health Administration patients (Chapter 2). One limitation of this prior

study was that a heterogeneous cohort of patients with bipolar disorder, depression or other severe mental illness was examined, with the intention of maximizing statistical power to observe associations over a uniform, one-year period after treatment initiation. This study focuses the examination of suicide risks only among individuals with a diagnosis of bipolar disorder, the principal psychiatric patient group receiving lithium or valproate.

Similar to our prior study (Chapter 2), a nationwide cohort study of lithium and valproate recipients was conducted that was designed to approximate some aspects of randomized trials by: 1) matching patients based on a high-dimensional propensity score[42] and 2) deriving intent-to-treat risk estimates.[48] High-dimensional propensity score matching allows treatment groups hundreds of measured covariates to be closely balanced between treatment groups (similar to a trial). Intent-to-treat estimates also ensure that any differences in the risks arising after treatment discontinuation are included when evaluation the comparative effectiveness of two treatments. Intent-to-treat estimates also removes confounding arising after treatment initiation from differential selection of patients to continue or discontinuation treatment, although bias can still arise if treatment modification is present.

Using these methods we evaluated whether suicide risk differed between patients with bipolar disorder initiated on lithium compared to valproate.

Methods

Data Sources

Demographic characteristics, inpatient and outpatient mental and non-mental health treatment records, and outpatient pharmacy prescription data was obtained from the VHA National Psychosis Registry (a linked, de-identified healthcare databases of all VHA patients with at least one psychotic or bipolar disorder diagnosis). This study was approved by the Institutional Review Boards of the Bedford and Ann Arbor Veterans Affairs Medical Centers.

Study Cohort

Incident users (≥ 6 months of no lithium or valproate use) with recent VHA utilization receiving at least one outpatient prescription for lithium or valproate from April 1999 to December 2008 were identified having a diagnosis of bipolar disorder (Bipolar I, II, or NOS) within 30 days prior to medication initiation. Patients were excluded if they possessed potential nonpsychiatric indications for valproate or lithium, initiated mood stabilizer on an "as needed" basis, or both mood stabilizers simultaneously.

Exposure Determination

Receipt of lithium or valproate was determined by outpatient prescription fills. For the intent-to-treat analysis, all patients filling an initial outpatient prescription were followed until end of follow-up or death. Secondary analyses stratified follow-up time by whether patients were still receiving initial treatment. Patients were considered "as-

treated" until a ≥ 15-day gap occurred between outpatient prescriptions (adjusting for early refills) or upon initiation of the other medication. "Former users" consisted of patients after initial treatment discontinuation until follow-up time was censored upon treatment resumption or switching, death, or end of follow-up. This nonexposed "former user" group has been proposed as a potential index of residual confounding and/or selection during treatment under certain conditions,[48, 49] however, it most accurately represents the combined effects of residual confounding, selection during treatment, persistence of any active treatment effects, and discontinuation-associated risks (e.g., rebound mania, anxiety, etc.).

Outcome

Date and cause of death was obtained from National Death Index files for 1999-2009.[56]

Propensity Score Modeling

Over 900 covariates, including a large number of potential psychiatric and nonpsychiatric risk factors for suicide from numerous administrative data domains, were derived from VHA databases (Chapter 2). Our approach generally followed the "high-dimensional" propensity score method,[41, 58] except variables were individually constructed and evaluated, instead of using automated variable construction and selection procedures.

Statistical Methods

The propensity score was calculated using logistic regression. Patients initiating lithium and valproate were then 1:1-matched using calipers of 0.2 standard deviations of the propensity score logit,[60, 61] resulting in 97.4% matching of lithium-initiated patients. Balance in covariates between the treatment groups was assessed using standardized differences *(*equivalent to Cohen's d effect sizes, with a difference of ≥ 0.10 considered as indicating significant imbalance).[42]

Statistical significance was determined using analytic methods reflecting matching (stratified Cox regression with sandwich variance estimators), except for analyses of "former users" (for which non-stratified Cox regression was used, since matching was not preserved).[61] The primary analysis examined intent-to-treat associations between treatments over 0-365 days. Secondary endpoints examining intent-to-treat treatment-risk associations over 0-90 days and 0-180 days. All analyses were performed using SAS, version 9.3, except the calculation of standardized differences, performed using Microsoft Excel 2010.

Secondary analyses

In addition to the secondary time periods examined for the intent-to-treat analyses, other secondary analyses were also conducted. The association between the treatments and suicide risk was examined among patients who were still receiving their initial treatment over 0-90 days, 0-180 days, or 0-365 days. Follow-up time for these "as-initially-treated" analyses was censored once ≥ 15 day had transpired from the end date

of their most recent prescription of the initiated mood stabilizer (lithium or valproate), at which point they were considered to have stopped treatment. Follow-up time was also censored immediately if they received a prescription for the other mood stabilizer, if they died from nonsuicide causes, at the end of follow-up. The association between initiation of lithium and valproate and suicide risk was also evaluated for patients who had stopped their initial treatment. Follow-up time for these "former users" analyses commenced on the 15th day after their last prescription fill of their initial treatment, and was censored immediately upon either resumption of the initiated treatment or switching to the other treatment, death from other causes, or the end of follow-up.

In addition, we also compared the suicide risk associated with the treatment groups prior to matching. Finally, the prevalence of diagnostically-coded suicidal ideation (V62.84, a code for suicidal ideation only available from the years 2005-2008) in the 30 days prior to treatment initiation was compared between the treatment groups among the patients for whom this information was available (< 50% of the full sample).

Results

The two treatment groups were highly similar in virtually all measured covariates even prior to matching. Only 7 of 917 (0.8%) covariates had an initial standardized difference (i.e., prior to matching) of ≥ 0.10, including only three covariates directly related to mental health (Table 4-1). The largest initial standardized difference was only 0.17. Table 4-1 demonstrates how the hdPS-matching functions to bring covariates, even

Table 4-1. Baseline Covariates by Treatment, Before and After Propensity-Score Matching

Patient Characteristic	Unmatched Sample (Prior to Matching)					Matched Sample (After Matching)				
	Lithium		Valproate		Standardized	Lithium		Valproate		Standardized
	N	%	N	%	Difference	N	%	N	%	Difference
COVARIATES WITH SUBSTANTIAL INITIAL IMBALANCES[a]										
Current Other Mood Stabilizer	1934	16.2	3062	10.9	0.156	1799	15.4	1792	15.4	0.002
History of past Mood Stabilizer	4770	39.9	9279	32.9	0.145	4576	39.3	4616	39.6	-0.007
State-level Suicide Risk, quintile 4	1649	13.8	2948	10.5	0.102	1555	13.35	1542	13.24	0.003
Liver Disease, Mild	1049	8.8	1306	4.63	0. 166	898	7.7	896	7.7	0.001
Current Thiazide Diuretic	732	6.1	2765	9.8	-0.137	719	6.2	731	6.3	-0.004
Current ACE Inhibitor	1404	11.7	4379	15.5	-0.111	1380	11.9	1364	11.7	0.004
Current NSAID	1846	15.4	5504	19.5	-0.108	1812	15.6	1848	15.9	-0.008
SELECT ADDITIONAL COVARIATES OF INTEREST										
Age, 65-79 yo[b]	789	6.59	2173	7.71	-0.043	778	6.68	778	6.68	0
Age, 80+ yo[b]	102	0.9	251	0.9	-0.004	99	0.8	99	0.8	0
Sex (Female)	1766	14.8	3374	12	0.082	1687	14.5	1710	14.70	-0.006
Post-Traumatic Stress Disorder (PTSD)	2305	19.3	6442	22.8	-0.088	2250	19.3	2266	19.5	-0.004

Table 4-1. (continued)

Depression NOS	345	2.9	1066	3.8	-0.050	338	2.9	345	3.0	-0.004
Alcohol Dependence	2625	21.9	6619	23.5	-0.037	2537	21.8	2625	22.5	-0.018
Psychiatric Hospitalization, past 7 days	1419	11.9	4226	15.0	-0.092	1391	11.9	1374	11.8	0.005
Psychiatric Hospitalization, last year	9149	76.5	20565	72.9	0.081	8916	76.5	8933	76.7	-0.003
Current SSRI	3643	30.4	9584	33.9	-0.074	3553	30.5	3549	30.5	0.001
Current SNRI	709	5.9	1575	5.6	0.015	687	5.9	679	5.8	0.003
Current Olanzapine	1118	9.3	3009	10.7	-0.043	1098	9.4	1062	9.1	0.011
Current Clozapine	6	0.05	10	0.04	0.001	6	0.05	6	0.05	0.000
Recent Other Mood Stabilizer	1488	12.4	2854	10.1	0.073	1400	12.0	1404	12.1	-0.001
Any acute injury	2179	18.2	5238	18.6	-0.009	2081	17.9	2074	17.8	0.002
NonFatal Suicide Behavior (NFSB), NonMHdx, past30d[c]	15	0.13	52	0.18	-0.015	15	0.1	14	0.1	0.002
NFSB, MHdx, past 30d[c]	19	0.16	43	0.15	0.002	19	0.2	21	0.2	-0.004
NFSB, Outptdx, past 30d[c]	79	0.66	218	0.77	-0.013	75	0.6	77	0.7	-0.002

Table 4-1. (continued)

NFSB, NonMHdx, past 31-180 days[c]	18	0.15	27	0.10	0.016	16	0.1	16	0.1	0.000
NFSB, MHdx, past 31-180 days[c]	18	0.15	32	0.11	0.010	18	0.2	13	0.1	0.012
NFSB, Outptdx, past 31-180 days[c]	38	0.32	107	0.38	-0.011	38	0.3	39	0.3	-0.002

[a] Substantial Initial Imbalance refers to a standardized difference in prevalence between the lithium and valproate treatment groups of ≥ +/- 0.10.

[b] Age is presented in this format to simplify presentation. In actuality, age was modeled with 11 indicators.

[c] NFSB= Non-fatal suicide behavior; NonMHdx= Diagnosis code entered during nonmental health hospitalization; MHdx= Diagnosis code entered during a mental health hospitalization; Outptdx= Diagnosis code entered during an outpatient encounter.

when initial imbalances are modest, into closer balance in the matched sample (often into extremely close balance). After matching, every one of the 917 covariates was closely balanced between treatment groups to a standardized difference of ≤0.024 (and >87% of the covariates within a standardized difference of 0.001).

The intent-to-treat survival curve for our matched cohort is presented in Figure 4-1.

Table 4-2 provides the results from our primary, intent-to-treat analysis. Significantly increased hazards of suicide death associated with initiation of lithium treatment were observed over 0-365 days (HR = 1.50, 95% CI 1.05, 2.15, p = 0.026). In addition, when secondary endpoints of 0-90 days and 0-180 days were examined, significantly increased hazards of suicide death were also observed at 180 days (HR = 1.87, 95% CI 1.17, 2.97, p = 0.008).

Table 4-3 provides hazard ratios for our secondary "as-initially-treated" and "former users" analyses. No significant differences were observed in suicide risk among patients during the time in which they received initial lithium or valproate treatment. In contrast, substantially elevated risk were observed among former users of lithium at 180 days (HR = 6.10, 95% CI 1.37, 27.2, p = 0.018). This association was not significant over 0-365 days, results were no longer significant at 365 days (HR = 2.05, 95% CI 0.88, 4.79, p = 0.098) although a p value <0.10 was observed.

Table 4-4 presents the prevalence of diagnostically-coded suicidal ideation between treatments for the 10,608 patients for whom this data was available (patients

Figure 4-1. Survival Curves, 0-365 days, Intent-to-Treat Cohort

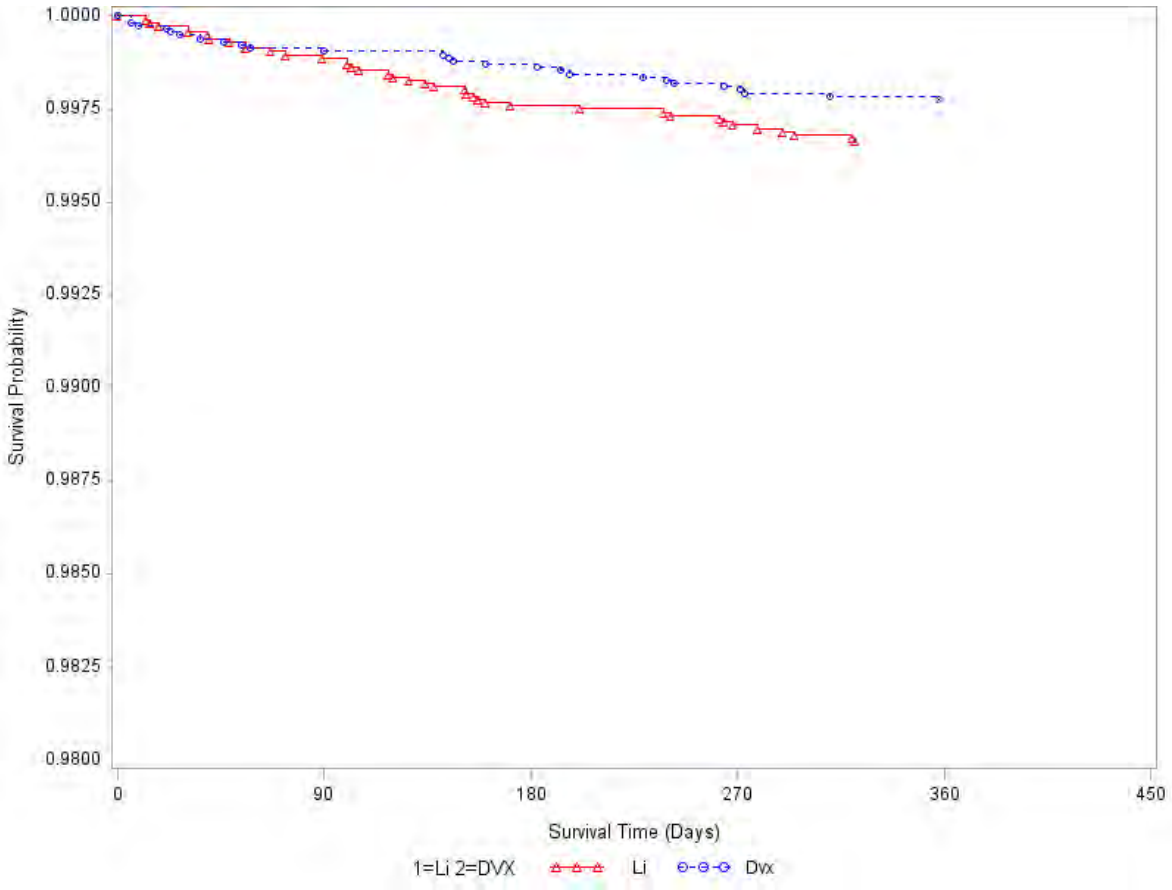


Table 4-2. Risk of Suicide Death by Treatment (Intent-to-Treat Cohort)

Length of Follow-up	Number of Suicides: Lithium	Number of Suicides: Valproate	Intent-to-Treat Hazard Ratio [Lithium/Valproate] (95% Confidence Interval)
0-90 days	13	11	1.18 (0.67-2.09)
0-180 days	28	15	1.87[a] (1.17- 2.97)
0-365 days	39	26	1.50[b] (1.05-2.15)

[a] p=0.008
[b] p=0.026

Table 4-3. Risk of Suicide Death by Treatment and Treatment Status (As-Initially-Treated versus Former User)

Length of Follow-up	As-Initially-Treated			Former User		
	Number of Suicides		Stratified Hazard Ratio (Li/Val) (95% CI)	Number of Suicides		Hazard Ratio (Li/Val) (95% CI)
	Li	VAL		Li	VAL	
0-90 Days	12	10	1.0 (0.5-2.0)	1	1	0.92 (0.06-14.7)
0-180 Days	13	11	1.0 (0.52-1.92)	12	2	6.10 (1.37-27.3)[a]
0-365 Days	15	15	1.0 (0.52-1.92)	16	8	2.05 (0.88-4.79)[b]

[a] p=0.018

[b] p=0.098

Table 4-4. Presence of V-code (62.84) denoting Suicidal Ideation in the 30 days prior to Lithium or Valproate Initiation[a]

Patient Characteristic	Patients Initiating Lithium	Patients Initiating Valproate	Odds Ratio (95% CI)
Suicidal Ideation	176	122	1.47[a] (1.16-1.86)
No Suicidal Ideation	5103	5207	

[a] for the portion of the cohort initiating treatment in 2005-2008

[b] p=0.0012

initiating treatment from 2005-2008). A diagnosis code for suicidal ideation had been entered for the patients initiated on lithium than valproate significantly more often in the 30 days prior to initiation of treatment (OR = 1.47, 95% CI 1.16, 1.86, p = 0.001).

Discussion

In this extensively-matched nationwide cohort of 23, 298 Veterans Health Administration patients with bipolar disorder initiated on lithium or valproate, initiation of lithium treatment was associated with increased risk of suicide at 180 and 365 days in intent-to-treat analyses. The high rates of treatment impersistence observed, however, means that the majority of days in the intent-to-treat estimate related to the period after initial treatment with lithium or valproate has ceased. As a result, the intent-to-treat results can be expected to reflect residual confounding to a greater extent than usual.

Furthermore, our secondary analyses indicate that some or most of the association between lithium initiation and increased suicide risk appears to be explained by substantially increased risk for suicide in patients who discontinued lithium compared to valproate. In contrast, active treatment with lithium was not significantly associated with increased or decreased suicide risk. Two important conclusions suggest themselves when these results are considered as a whole. First, the possibility that initiating lithium may pose a suicide hazard to some patients who discontinue it even within the first year of treatment needs increasingly thorough research and clinical consideration. The occurrence of such risks upon discontinuation is consistent with past findings indicating that the period post-lithium discontinuation may pose distinct suicide, suicidal behavior,

or mood episode risks. Second, the conclusion cannot be reached that active lithium treatment was not associated with reduced risks of suicide in this patient cohort, potentially even by a substantial, clinically-meaningful amount. Residual confounding in this study appears to bias against finding protective associations between lithium treatment and suicide, thus weakening any inferences concerning the presence or size of a protective association between active lithium treatment and suicide risks. Making further progress regarding possible confounding, however, will likely require either different study designs or different patient cohorts with lesser confounding.

These two conclusions regarding possible discontinuation-associated risks and residual confounding biasing against lithium were also the principal conclusions of our earlier study (Chapter 2). This earlier study included most of the bipolar individuals examined here along with patients with other mood and psychiatric disorders who were receiving lithium or valproate. However, these conclusions are reinforced and strengthened by these specific results, which show directionally similar, although even more extreme, patterns of risk over the study period.

The judgments concerning residual confounding also are reinforced by the apparent greater degree of channeling of patients with bipolar disorder with recently-diagnosed suicidal ideation to preferentially receive lithium (OR=1.47) than for the cohort as a whole (OR=1.30) (Chapter 2). Finally, the substantially increased risks of suicide observed at 180 days have even more of an "emergent quality" in this study (i.e., a greater change in the risk among former users than observed over 0-90 days) than the prior study of the whole cohort (Chapter 2). Although some caution is warranted since

the confidence intervals for the estimates of risk in patients after discontinuing treatment over 0-90 days and 0-180 days overlap, this suggestion of emergent risks are more consistent with risks that are associated with discontinuation (an event that takes time to occur), and is less consistent with confounding (at least with fixed confounding or with time-varying confounding that is most substantial close to initiation, a pattern suggested by some other suicide studies examining psychiatric medications.[17, 122]

For these reasons we conclude that our findings in this study simultaneously suggest that some degree of residual confounding biasing against lithium likely exists, as well as some degree of greater suicide risks upon lithium, compared to valproate, discontinuation.

The possibility of risks associated with lithium discontinuation appears consistent with several decades of research. Since the early 1980s evidence has accumulated that discontinuation of lithium places patients at increased risk for mood episodes. Subsequent work has demonstrated that the time to mood episode relapse can be as brief as approximately two weeks (more precisely, 13-19 days)[123] to 1-4 months,[79, 107] and that associated risks are especially pronounced if the discontinuation is abrupt.[79] One study reported that suicide risks were particularly prominent within only 30 days after discontinuing mood stabilizers, without specifying the specific risks observed with lithium, valproate, or carbamazepine discontinuation.[16] To our knowledge, however, no study has systematically examined risk from discontinuation occurring shortly after medication initiation, a period in which patients may still be at heightened vulnerability for suicide in general.

Because of these findings consistent with possible discontinuation-associated risks, this study suggests a clear need exists for clinical systems and providers to encourage patients to continue with their lithium treatment once initiated. Educational/psychosocial interventions have been developed to promote adherence with treatment.[116 117, 118] Clinical systems and providers also may need to take steps to minimize suicide risk in patients who stop lithium treatment, whether at their own initiative or as directed by their provider. Available evidence suggests that when discontinuation occurs, it should be done gradually if possible.[79] When feasible, increased clinical monitoring also should occur after discontinuation, as has been previously recommended for mood stabilizers to minimize risk of mood episode relapse.[78]

Further assessment and/or research also may be needed concerning the more difficult question of whether lithium should be avoided in certain patients for whom risks of discontinuation appear high or adherence to follow-up monitoring likely to be low. A recommendation not to initiation lithium in patients deemed likely to discontinue the treatment has been made 20 years, although it was based on the very limited randomized evidence available at that time.[124] Finally, the fact that significantly increased intent-to-treat risks are observed, despite the fact that some of these risks may result from confounding and that the risks appear associated with lithium discontinuation rather than during lithium treatment, should encourage added caution on the part of clinical systems considering whether to promote expanded use of lithium as a possible suicide preventative.

It is important to recognize that our findings do not necessarily suggest a lack of effectiveness for lithium against suicide during active treatment. Residual confounding biasing against lithium would imply that the central HR of 1.0 observed for active lithium treatment may actually be a composite of some degree of protective association from lithium coupled with some degree of residual confounding in the opposite direction (biasing against lithium). This bias against lithium is suggested both the suicidal ideation findings (more diagnoses in patients initiating lithium than valproate) and by comparison of matched versus unmatched effect estimates, which indicated greater risks associated with lithium treatment prior to matching (data not shown). Furthermore, propensity score methods also may amplify to some degree whatever residual confounding exists after control of the measured covariates, and some degree of unmeasured confounding is plausible in our analyses. Any such amplification is likely modest, however, given the low c statistic.

A second reason exists to view our study as not necessarily indicating a lack of effectiveness for lithium against suicide during active treatment. Both our study design and the very high rates of treatment discontinuation, makes it difficult to infer about suicide risk after 365 days (although we present some very limited data for the overall cohort in Chapter 3). Some studies have specifically reported that treatment with lithium for >1 year was required to observe significant reductions in suicide risk.[67-69]

Nevertheless, our results appear generally consistent with two recent randomized trials that had durations of 2-2.5 years. Neither of these studies observed statistically significant reductions in suicide behavior amongst patients initiated on lithium. It is

worth nothing, however, that both studies observed nonsignificant findings of effects that still were in the direction of reduced suicide risk. The findings from these studies have been combined in a recent meta-analysis which derived a substantial, bust still nonsignificant, reduction in was associated with a nonsignificant reduction in nonfatal suicidal behavior compared to valproate (OR = 0.64, 95% CI 0.30, 1.36, p = 0.24). Treatment persistence was substantially lower in this study than those trials, perhaps contributing (given the likely direction of confounding) to the different direction of association observed in our study.

The complications engendered by the possibility of residual confounding suggests that further research should either strongly consider study designs that can address unmeasured confounding (randomized trials and possibly instrument variable analyses) or perhaps focus on cohorts in which lithium is more universally used, such as some areas of Europe.

This study has additional limitations. These include the fact that information was available for VHA outpatient prescriptions only, serum medication levels were not available, suicide events were relatively few, and that there was an inability to control for some relevant confounders such as suicidal ideation, plans, and means, mental health symptoms, psychosocial stressors, and outpatient hospitalizations. In addition, our design choice of not restricting the cohort to patients receiving monotherapy with mood stabilizers (lithium and valproate) helped preserve power and generalizability of findings, but would be expected to add uncertainties. In particular, changes in other psychiatric medications over the one year of follow-up were not included in the analysis. However,

the patient cohorts were closely balanced on a very wide variety of psychiatric medications or medication classes at study initiation. Confounding amplification is also possible, although likely was fairly minimal given the very modest c statistic. Nevertheless, the potential for this study design to both remove more measured confounding than some other study designs, but also amplify residual confounding, should be kept in mind since several plausible confounders were not able to be included.

Finally, generalizability of this sample outside other Veteran samples with similar sex distribution (predominantly male) and rates of Post-Traumatic Stress Disorder (PTSD) and substance use is uncertain. In addition, generalizability of the intent- to-treat estimates is further limited by the fact they are most applicable to cohorts with similar rates of treatment discontinuation.

Conclusions

In this study of US Veterans with bipolar disorder, initiation of lithium, compared to initiation of valproate, was associated with significantly higher suicide risks over 0-180 and 0-365 days in our primary, intent-to-treat analyses. These analyses included the outcomes of all patients initiating treatment were considered regardless of whether they occurred during active treatment or after discontinuation any subsequent treatment discontinuation (>90% of patients discontinued had their treatment by 365 days). Significantly increased suicide risks were observed in secondary analyses associated with patients who had discontinued lithium, compared to valproate, treatment over 0-180 days, but not among patients receiving active lithium treatment. Several lines of evidence

support the possibility of some residual confounding biasing against lithium; nonetheless, our findings suggests that clinical systems and providers should consider taking steps to minimize treatment discontinuation once lithium is started and increase the monitoring of patients after lithium discontinuation. In addition, the likely presence of some residual confounding biasing against lithium largely precludes judgments about whether lithium has effectiveness against suicide in this cohort. Given the lack of effective interventions against suicide, further research examining the balance of potential risks and benefits associated with the initiation of lithium treatment is clearly needed.

CHAPTER V

Final Summary and Conclusions

This set of studies has examined several important aspects of the question of whether the mood stabilizer lithium might serves as an effective suicide preventative in United States Veterans Health Administration (VHA) patients. Lithium's effectiveness for suicide prevention was examined among incident psychiatric users overall, as well as specifically among patients with bipolar disorder. In addition, the question of whether any potential benefits of lithium versus suicide might be counterbalanced by increased mortality for other causes of death was also addressed. Lithium's association with both suicide and nonsuicide mortality has been examined in comparison to a popular and more commonly-used alternative treatment, valproate.

***

In mental health research, nonrandomized comparative effectiveness research may be particularly needed to address questions regarding the association of psychiatric medications with irreversible or terminal endpoints. Randomized trials examining such endpoints, at least as primary outcomes, are likely to either be judged unethical or unpopular with patients. Some limited data regarding suicide and nonsuicide mortality does exists from randomized trials and is likely to continue to be accrued, but acquired and reported as "serious adverse events" rather than as study primary outcomes. Our

study employed high-dimensional propensity scores, a recent innovation in nonrandomized comparative effectiveness studies, but one for which considerable validation data has been acquired.[41, 125, 126] Furthermore, our studies mimicked randomized trials in a limited degree in producing groups of patients receiving either medication that were highly similar on numerous covariates and also through the use of intent-to-treat effect estimates.

Unfortunately, nonrandomized studies in mental health can be expected to be particularly challenging. The likelihood of confounding, especially confounding due to unmeasured confounders, is especially high and difficult to predict. Propensity score methods are able to balance measured factors included in the propensity model, but not unmeasured factors. So many aspects central to mental health care are not routinely measured in administrative clinical datasets, or even recorded in chart documentation. The latter consideration means that even labor-intensive chart review is still unlikely to fully measure the factors that may contribute to confounding. These unmeasured or incompletely measured elements include, for example, such important factors as outside hospitalizations, recent mental health symptoms and psychosocial stressors. In addition, certain "measured covariates" in mental health may be very imperfectly measured or modelled, such as substance abuse. Although diagnostic categories exist which differentiate substance abuse from substance dependence, for example, these categories almost certainly are highly imprecise measures of what is of most interest: the level of substance use or misuse occurring right at the time of treatment initiation.

As a result of these uncertainties, additional caution must be exercised in the interpretation of nonrandomized mental health treatment studies compared to other nonrandomized treatment studies. The possibility of the residual confounding must always be a prominent consideration in the interpretation of nonrandomized mental health treatment studies.

For certain outcomes, however, that are often not examined in mental health research, such as nonsuicide mortality, reasonable assumptions can be made that at least some direct forms of confounding are likely to be less prominent. "Confounding by indication" refers to the tendency for providers to choose different treatments based on the patient's condition, and can be alternatively termed, most briefly, as "channeling," or perhaps most descriptively as "confounding by provider intention." "Confounding by indication," is specifically likely to be reduced in studies of non-indicated uses or ancillary effects of mental health medications (such as their associations with nonsuicide mortality) compared to the confounding by indication that would exist for studies examining indicated uses of the medication. Mental health prescribers are not tasked with primary responsibility for addressing a patient's physical health. Some degree of confounding by indication in studies of psychiatric medications and nonsuicide mortality, however, remains still feasible. First, mental health providers do not ignore the physical health condition of their patients, it is simply not their *primary* focus. Second, medical contraindications exist for mental health treatments (e.g., kidney- and cardiac-related contraindications for lithium, and liver or bone marrow- related contraindications for valproate), that plausibly could affect nonsuicide mortality risk. Third, to the extent that

providers (or patients) believe one of the two medications to be more effective, then this medication might be preferred for those patients who are most severely mentally ill. To the extent that severity of mental illness may impact mortality risk, this prescribing tendency, regardless of the actual effectiveness of the medication, would be expected to lead to some degree of nonsuicide mortality "confounding by indication."

***

Methodological Conclusions

Part of the value of this study has been as a "testing ground" to determine whether a recent innovation, high-dimensional propensity scores, can be profitably applied the VHA's large, extensive databases. This set of studies has gone beyond a simple application of high-dimensional propensity score studies to integrate other innovations, such as the use of intent-to-treat, current user, and former user risks in an effort to enhance the inferences that can be made from nonrandomized studies. This set of studies has demonstrated that:

1) *Implementation of a high-dimensional propensity score is feasible in VHA databases.* Our high-dimensional propensity score was not executed using the automated variable generation methods of the originally-proposed method. Yet, in approximately a year of half-time programmer effort we were able to generate almost 1,000 propensity score variables that not only well represented the detail available in VHA databases, but also had likely *a priori* relevance.[127] That is,

each variable chosen for inclusion was judged to have some non-negligible probability of being associated with mortality (either suicide or nonsuicide mortality). Our judgments were borne out somewhat by the fact that when variables were excluded from our propensity score that were empirically observed to have a < 20% association with outcome in univariate comparisons, less than half the variables were excluded.

In addition, we were able to identify opportunities to "semi-automate" certain important variables. Most notably, this included using the first 2 digits of the VHA's 3 digit "medication class codes" for generating our >100 nonmental health medication covariates based on current or recent but apparently discontinued use of medications. Not only did this approach speed variable construction but using a more aggregated form of important covariates has been recently associated with better performance [128].

2) *Any confounding amplification is likely to be modest.* Although concerns about confounding amplification only began to be widely discussed after initiation of our project, through our consultation with Dr. M. Alan Brookhart we had already adopted an approach which appeared to keep any confounding amplification relatively modest. Our approach scrutinized in particular those variables with a particularly strong association with exposure to assess further whether they were plausible confounders. Of note, no variables were observed to meet Dr. Brookhart's original criteria of concern: an odds ratio of exposure of ≥ 4.0. (In

fact, no covariates were observed that had even an odds ratio of exposure of ≥ 3.0. This is another potential indicator that our two treatment groups may not have been extremely confounded even prior to the application of the propensity score). Therefore, we extended this scrutiny to covariates with an odds ratio of exposure between 2.0 and 3.0. This process led to the non-inclusion of a handful of variables that were judged to potentially likely not to have a genuine relationship with the outcome. Most importantly, our c statistics for our propensity score models were relatively modest (the highest c statistics being in the range of 0.69-0.70). Recent simulation has suggested that propensity score models with explanatory power in this range have modest confounding amplification in the range of ≤ 60%, although the precise specifics of this simulation was based on R^2, not c statistics.[47] The possibility of some degree of confounding amplification reinforces the need to consider a role for residual confounding in our study, in spite of the extensiveness of our efforts to deal with measured confounding.

3) *Former user risks appear to have some distinct value in the interpretation of findings from nonrandomized studies (and presumably from randomized studies as well), although their interpretation is complex*. The most extreme demonstration of this value occurred during the interpretation of the intent-to-treat suicide risk results for the individuals with bipolar disorder alone. Splitting the analysis into individuals receiving initial treatment and former users made it clear the entire association between lithium and increased risk related to risks among

patients who had stopped initial treatment. In addition, there appears to be value for interpretation in separating "former users" – those individuals who are no longer exposed to either medication being studied from "resumers" – patients who resume either treatment, since nonexposed individuals have value as potential indicators of residual confounding at baseline or arising during treatment.[48, 49]

It is interesting to note that depending on available sample size and number of outcomes, several additional categories of "former users" could be envisioned with progressively increasing stringency in their definition of "nonexposure." For instances, if sample size permits, it may prove valuable to examine a special class of former users which examines the follow-up time for individuals who have stopping their medication studied and not initiated *any other* psychiatric medication subsequent to this discontinuation. Even more rigorous would be a requirement that "former users" not include any individuals who had started other psychiatric medications on or after the date of initiation of the medication under study. In this case, the "former user" period would represent a period of nonexposure to any psychiatric medication which was not present at baseline. Given still greater sample size, the most rigorous definition would be to examine only patients initiating psychiatric medication monotherapy involving the study medications; thus, the former user period would consist of follow-up time free of exposure to *any* psychiatric medication. That design among other things, would prevent uncertainties arising from changes as seemingly as minor as possible dose increases of psychiatric medications being received concomitantly.

As the definition of "former user" becomes more restrictive, however (especially if a restriction to monotherapy is enforced), generalizability of the findings would become restricted.

4) *Stratification appeared to play a useful role in facilitating interpretation of our results, even with our limited power.* The clearest demonstration of the value of stratification arises by stratification of the follow-up period into multiple time periods (0-90 days, 0-180 days, and 0-365 days for all 3 studies, and, for the first suicide risk analysis, division into the mutually exclusive categories of 0-90, 91—180, and 181-365 days as well). In addition, our stratification of suicide risk by treatment by psychiatric diagnoses (bipolar disorder versus nonbipolar disorder) was also valuable. The stratification of follow-up time permitted the detection of significantly elevated risks in former users over the 0-180 day period in all 3 analyses which would have been missed with simply an examination of 0-365 days. Detection of this elevated risk had particular value in the nonsuicide mortality manuscript, since it occurred in the opposite direction of the significant associations of reduced nonsuicide mortality with lithium initiation in the intent-to-treat analyses (0-90 days), and with still receiving lithium treatment in the current user analyses (all 3 time periods). This suggests that simple confounding (i.e., in the same consistent direction), even if it varied in size, could not explain the entirety of both the 0-90 day findings suggesting potential effectiveness for

lithium and the 0-180 day former user findings suggesting possible distinct risks upon discontinuation.

Our stratification by diagnoses yielded the observation that in the bipolar cohort both risks in former users and intent-to-treat risks were significant, and in the hazardous direction. This suggests that the elevated risk in former users arises from either baseline confounding (which would then also be expected to be biasing the treatment effect estimates in the intent-to-treat and still receiving initial treatment cohorts) or from risks associated with discontinuation of the treatment, rather than from confounding arising after treatment initiation from selection occurring during follow-up (at least pertaining to risk for from non-medication related outcomes, which would not affect ITT estimates).

5) *Examination of non-included factors can help provide insight into the likely direction of residual confounding. For the sui*cide mortality analyses, a potentially important covariate was likely present for approximately half the sample: suicidal ideation, as reflected by a diagnostic code (V62.84). This covariate likely only represent a fraction of reported suicidal ideation (it was present in <2% of our sample in the 30 days prior to medication initiation), and the fact the code only came into existence in 2005 meant it could not be present for the approximately half of patients initiating medication prior to that time. Yet, examining this covariate helped reinforce the possibility that some residual confounding biasing against lithium likely existed in our suicide risk studies: for

both studies, significantly more patients initiating lithium had a V-code denoting suicidal ideation in the past 30 days than patients initiating valproate. Not only might this imbalance account for some amount of residual confounding directly, this residual imbalance after propensity score matching potentially suggests that a similar imbalance may occur in other suicide-related risk factors, although this cannot be concluded with certainty.

***

Comparative Effectiveness Conclusions

This set of studies comparing lithium and valproate had several important and unexpected results. First, at the most basic level, the high rate of treatment discontinuation we observed in this VHA sample poses a substantial practical barrier to any efforts to employ lithium for the purposes of influencing either suicide or nonsuicide mortality. Approximately 75% of patients had discontinued (or been discontinued by their providers) their lithium or valproate treatment (at least temporarily) within 6 months of initiating treatment. Our observed rates of treatment discontinuation were surprising to us for a medication intended to be a long-term treatment of a generally chronic condition. However, the rates observed in this VHA sample were not very dissimilar from those observed in the only two other similar studies (i.e., studies examining broad samples of incident users) that we could identify.[113, 114]

More importantly however, were the findings that these high rates of treatment discontinuation were coupled with intent-to-treat associations with suicide risk that either

favored lithium to only a fairly minimal, nonsignificant extent (for the cohort as a whole), or that significantly favored valproate (for individuals with bipolar disorder). These intent-to-treat results suggest that it is possible lithium initiation among patients with bipolar disorder, at least over the first year of treatment, poses a net suicide hazard. This is of particular concern since individuals with bipolar disorder are the core target group for psychiatric treatment with lithium. Further examination of the results from individuals with bipolar disorders revealed that the significantly increase intent-to-treat suicide risks were associated almost exclusively increased suicide risks associated with individuals who had discontinued their initial lithium treatment. The extent to which this increased risk after lithium discontinuation relates to residual confounding biasing against lithium, or risks produced upon lithium discontinuation that exceed risks produced upon valproate discontinuation, however, is unclear.

Our intent-to-treat results are compatible, in the general sense of failing to establish a benefit for lithium in suicide prevention, with the only trial explicitly designed to compare lithium and valproate for suicidal behavior prevention.[53] This study found no significance differences between lithium and valproate, and only modest nonsignificant differences, although the study was very small (49 patients in each arm).[53]

The observation within all three studies that for at least one time period the risks associated with lithium discontinuation were significantly greater than risks associated with valproate discontinuation may be consistent with findings from almost 3 decades of research of distinctly increased risks for mood episode relapse shortly after lithium discontinuation. Of note, as early as 1994 one commentator interpreted randomized

research reporting intent-to-treat outcomes as indicating that the risks of mood episode relapse exceeded the benefits of mood stabilization from active lithium treatment unless patients remained on lithium for at least two years.[124]

Some studies have also reported increased risks of suicide upon lithium discontinuation. To our knowledge, however, risks associated with discontinuation have not been previously reported to occur in relation to lithium discontinuation occurring after only as brief a course of treatment as examined here (≤ 180 days). At a minimum, the set of studies described here puts the issue of potential discontinuation-associated risks front and center in assessments of whether to encourage the initiation of patients on lithium as a suicide preventative. It also creates a need to consider strategies to minimize discontinuation, educate patients and providers about the risks of discontinuation, and to monitor patients after discontinuation.[1]

The association of significantly lower nonsuicide mortality risks associated with lithium than valproate among patients still receiving initial treatment (and among the

---

[1] Not explored extensively here is the additional implication from these findings that, to the extent that lithium has effectiveness against suicide, it is possible that this effectiveness will be greater in populations without bipolar disorder than those with bipolar disorder, since counterbalancing risks upon discontinuation may not be prominent. This conclusion is suggested by the fact that the findings in the cohort with bipolar disorder are suggest lesser benefits and/or greater harms associated with lithium treatment than the findings for the cohort as a whole. By implication, this suggests that some degree of greater benefits and/or lesser harms pertaining to suicide risks are likely to be associated with lithium treatment in individuals without bipolar disorder.

entire intent-to-treat sample of patients initiating lithium over the first 90 days of treatment), in combination with the significantly higher nonsuicide mortality risks observed among patients stopping lithium over 0-180 days, suggests that some degree of genuine nonsuicide mortality risk difference exists between lithium and valproate. The overall direction of the association (i.e., whether lithium initiation is beneficial or detrimental) is not clear, and difficult to determine since such judgments depend on the level of unmeasured confounding.

In addition, two primary insights concerning likely confounding also suggest themselves. First, our studies suggested that the comparison between lithium and valproate initiators was relatively unconfounded by measured covariates. In all the studies, only a handful of the >900 covariates in the propensity score differed by a degree typically seen as indicating substantial imbalance (i.e., a standardized difference of ≥ 0.1, a measure equivalent to the better-known Cohen's *d* statistic). For instance, in the bipolar disorder cohort, < 2% of covariates had an initial imbalance between the treatment groups of a standardized difference of ≤ 0.1. This suggests that many types of information available clinically to providers did not appear to influence treatment choice (for instance, diagnoses, or the presence of a recent mental health or nonmental health hospitalizations). It also is consistent with, although does not prove, the possibility that providers are substantially relying on their own preferences, based perhaps on training, marketing, or prior experience when recommending lithium or valproate treatment, rather than the patient's condition. If so, this circumstance would be beneficial for the purposes of a nonrandomized study, since it implies baseline confounding may be fairly minimal.

An apparent "provider preference" has been observed to be a major influence on prescription choice for other mental health medications and been exploited in certain circumstances as a candidate instrumental variable.[129]

Second, despite our approach to minimize confounding amplification, and evidence suggesting initial overall confounding from measured factors may have been modest, the potential for at least some degree of residual confounding also appears substantial. For both the suicide and nonsuicide mortality analyses, the modest c statistic, while desirable to minimize confounding amplification, intrinsically suggests that a substantial portion of the variance in receipt of lithium or valproate remains unexplained in the propensity score model. Thus, while it is certainly true that much or most of this variance may relate to factors unrelated to outcome and thus not confounders (a prime example would be any provider preference for one medication over the other), the modest c statistic also implies ample opportunity exist for some predictors of exposure that are genuine confounders to have escaped modeling.

***

In conclusion, despite well-known limitations of nonrandomized research concerning control of confounding, and the high probability that these limitations are likely to be particularly germane to nonrandomized mental health treatment research, this set of studies has made significant contributions to the scientific literature. Through application of the high-dimensional propensity score, examination of intent-to-treat, current user, and former user risk, strategic use of stratification, and enhancement of the

analysis by examining important factors not able to be included in the propensity score, this set of studies has managed to arrive at several important conclusions. The most notable conclusion concerns the imperative emerging from all three studies (and reinforcing previous findings) to plan ways to minimize and manage any potential risk related with the discontinuation of lithium. A second prominent finding is that some degree of nonsuicide mortality risk difference may exist between lithium and valproate. Our nonsuicide mortality study therefore reinforces the need for the overall mortality impacts of psychiatric treatments to receive increased clinical and research attention. Finally, careful interpretation of the suicide risk studies reveals that although these studies do not directly support a sizeable benefit of active lithium treatment in reducing suicide risk, the possibility of a sizable benefit cannot be rigorously excluded. A degree of confounding biasing against lithium in the suicide studies appears likely and potentially is substantial. This set of studies therefore has produced a clear mandate for further research, and will hopefully inform multiple subsequent studies inside and outside the Veterans Health Administration.

APPENDICES

Appendix 1

(Additional information of particular relevance to Chapter 2)

Appendix 1-6. Modified Propensity Score Analysis

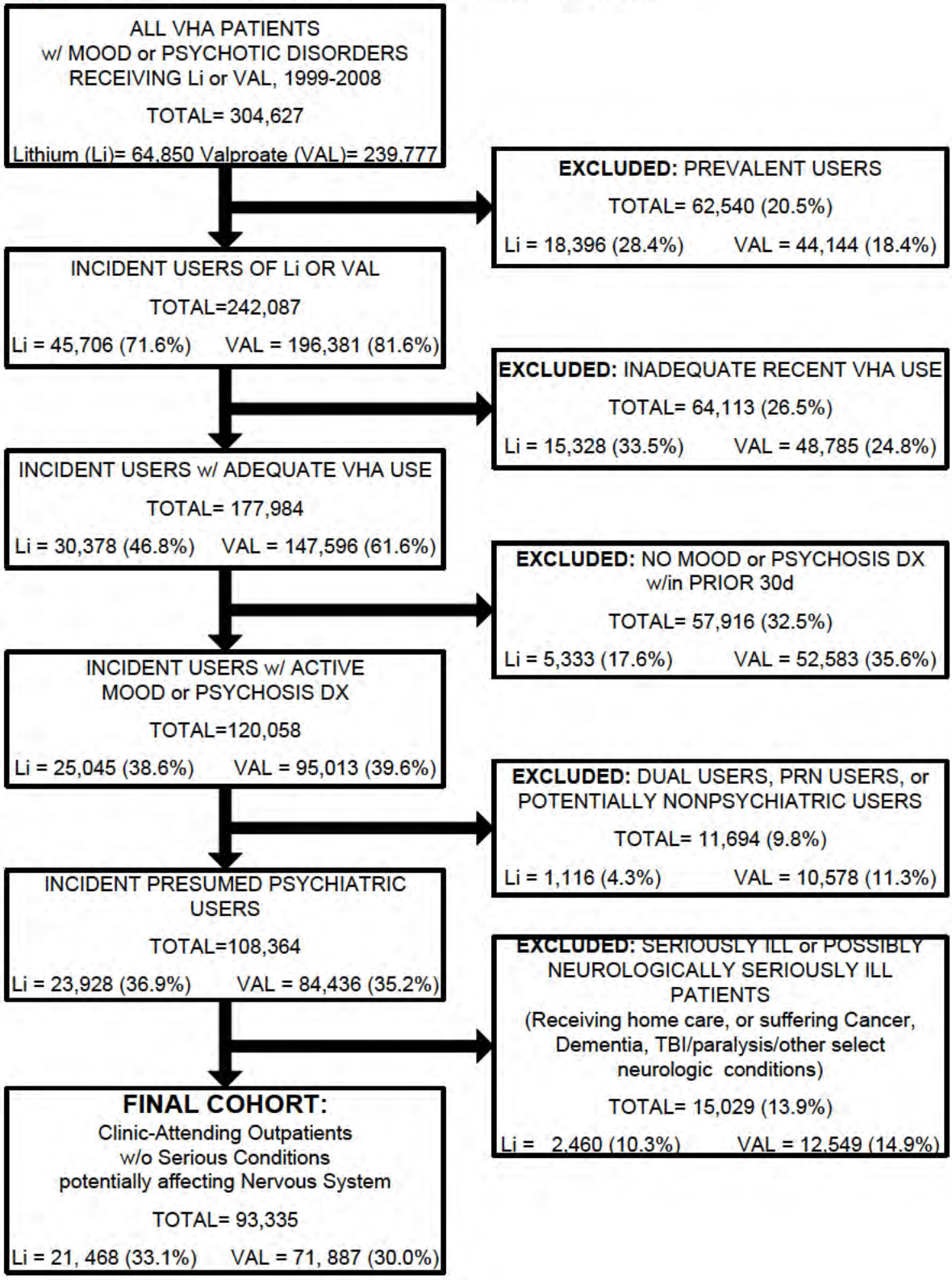
Appendix 1-7. Suggestions Concerning Future Research

Appendix 1-1. Diagnostic Codes Included in the Cohort

Since the databases used in this study were initially developed for use in tracking the care delivered to a broad collection of Veterans Health Administration (VHA) patients with depressive or psychotic disorders, a considerable range of diagnostic codes were available for inclusion during database construction. To maximize power and because existing literature suggested than any suicide benefits from lithium might span a variety of diagnoses,[51, 55] we decided to retain a broad group of eligible mood and psychotic diagnoses in the cohort. Virtually all cohort members had received a diagnosis of bipolar I, bipolar II, or bipolar NOS, depression NOS, major depressive disorder, schizophrenia, schizoaffective disorder, or "other psychoses" (including Psychosis NOS) within the past 30 days, and the prevalence of these diagnostic categories were highly similar between the two matched treatment groups (i.e., within a standardized difference of <0.018). Table 2-1 in Chapter 2 provides the data for the final prevalence for each treatment group of those diagnostic categories with initial substantial imbalances between the treatment groups.

Specifically, patients could enter the cohort with receipt of at least one of a number of ICD-9 codes in the past 30 days prior to lithium or valproate initiation. The most common codes by far were 296.0-296.99 and 311. Much less common were 295.0-295.9, 297.0-297.3, 297.8-297.9, 298.0-298.4, 298.8, 300.4, 301.12, 309.0-309.1, and 293.83. Only a few diagnoses predominated: bipolar disorder, major depression, and depression not otherwise specified; for instance, as Manuscript Table 1 indicates, less than 6% of patients in both treatment groups had a diagnosis of schizophrenia or "other psychoses." In addition, (not shown in Table 2-1) approximately 5% had schizoaffective disorder. Thus, although our final cohort did include a few individuals with schizophrenia or "other psychoses," the final diagnostic composition consisted of only 11% of individuals with a psychotic disorder, and some of these individuals also had diagnoses of eligible mood disorders within the past 30 days. Furthermore, although our entry criteria did permit some increased diagnostic heterogeneity compared to past studies, the propensity score-matching did produce an extremely similar prevalence of each diagnosis within the two treatment groups (i.e., within a standardized difference of 0.018 for all diagnostic categories).

Appendix Figure 1-1. Flowchart of Study Cohort Derivation



Appendix 1-2. Additional Information Concerning Variables Included in the High-Dimensional Propensity Score

DEMOGRAPHICS AND YEAR OF ENTRY

Demographics: Indicator variables were used for age (< 35 years old, ≥ 80 years old, and intervening 5-year age intervals), sex, and race/ethnicity as recorded in VA system. (Race information is relevant to studies of suicide risk because suicide rates vary widely based on race. When information on race was missing it was imputed using methods previously developed). In addition, indicator variables were also included for marital status (single/married/separated or divorced/widowed), income, disability status (as indicated by percent of "service connection" of a particular disability), distance to Veterans Health Administration (VHA) facility, urban/rural location of the facility where they are obtaining care, and fiscal year of medication start.

UTILIZATION

Utilization variables are derived from VHA clinic stop codes, a set of approximately 500 codes used to categorize each outpatient encounter. These codes result in classifying care provided into considerably broader categories of care than CPT codes used in "high-dimensional" propensity scores ,[41] reducing the need to consider whether codes should be aggregated or whether information is lost without such aggregation.[127]

General Mental Health and NonMental Health Utilization: We calculated the total number of VHA clinic stop codes relating to encounters with providers over specific time periods. We then used indicator variables to indicate whether, and at what frequency mental health and nonmental health encounters had occurred over periods as brief as the last 7 days before medication initiation to longer time periods occurring over the previous two years.

For general mental health utilization, we also constructed variables reflecting the total number of hospitalizations (as indexed by discharge dates), and variables dividing total MH provider visits into four subtypes (diagnostic interviews, medical management visits, and individual and group psychotherapy visits) over different time periods. For general nonmental health utilization, we also included variable representing the number of nonmental health hospitalizations and the number of surgery clinic and specialist visits (based on stop codes) during particular time periods. Also, variables were constructed reflecting the total ER/Urgent care visits, lab visits, and presence and absence of a flu shot in the last year (one possible indicator of preventative care).

Lastly, for both general mental health and nonmental health utilization, we included indicator variables for the total number of mental health and nonmental health medications, divided into medications that people were receiving on the lithium/valproate start date, the number of medications that they had very recently been taking but for which an active prescription did not exist on the date of lithium/valproate start (termed "Possibly Discontinued"), and the number of medications recently received (within the

last 180 days) but not received in the last 30 days ("Recently Discontinued"). The types of medication considered "mental health" is described under the subsection "Medications" below and in Appendix 1-3. All other medication types were considered "nonmental health medications."

The distinction between "general/basic" utilization and more specific outpatient utilization is somewhat subjective. For instance, we included the total number of lab visits under "general utilization" but included number of X-Rays, EKGs, and other diagnostic tests under "Non-Mental Health Diagnostic Tests."

Mental Health and NonMental Health Outpatient Utilization: Clinic stop codes were classified with indicator variables to reflect whether a patient had attended no visits of that type, a single isolated visit, or repeated visits (2 or more visits of that type) within a time period. The two time periods examined were the last 180 days prior to lithium/valproate start, and the prior 181 to 365 days before lithium/valproate start. For mental health outpatient utilization, visits were classified as occurring with psychiatrists, psychotherapists, in the general mental health clinic, primary care behavioral health clinic, substance use disorder clinic, or Health Care for Homeless Veterans clinic, with additional indicators for visits involving group treatment.

A much greater variety of stop codes exists for nonmental health outpatient utilization. We chose all stop codes appearing for ≥ 5% of either treatment group in either the last 180 days or days 181 to 365 prior to medication start and other, lower prevalence clinic stop codes thought *a priori* to be of importance as indicating potentially substantially compromised physical health (e.g., pacemaker clinic, etc.).

In addition, nonmental health stop codes also were also used to construct the diagnostic testing module described below.

Mental Health and NonMental Health Hospitalizations: The VHA uses approximately 90 bedsection codes to classify hospitalizations by the type of care received. The 30 bedsections that relate to mental health hospitalizations were classified into 4 larger classes: Psychiatric-focused hospitalizations, Substance Abuse-focused, Residential/Day program, and Domiciliary Program (longer-term housing).

Because suicide risks with relation to mental hospitalization appear to be time-dependent, we focused on capturing timing of hospitalization and the nature of the most recent hospitalization. We constructed multiple indicators to reflect the timing of the latest discharge date relative to medication initiation, as well as characterizing that latest hospitalization into one of the 4 classes of mental health hospitalizations.

With regard to bedsection codes for NonMental health hospitalizations, a few codes were consolidated when counts were observed to be particularly low (e.g., dermatology bedsection discharges), but in most cases a simple indicator variable was developed to reflect either that the patient's most recent hospitalization had been of that bedsection type, or that any of their hospitalization bedsections in the two years prior to medication start had been of that bedsection type. These latter variables were constructed both as a measure of overall disease burden (of conditions of a severity requiring hospitalization), because for some progressive conditions earlier hospitalizations or

diagnoses can actually reflect worse health prognosis, [130] and because failing health is one risk factor for suicide. These variables included ICU bedsections, "Step Down" Bedsections, Telemetry Bedsections, General Medicine Bedsections, Specialty Medicine (e.g., Neurology, Cardiology) Bedsections, Surgery Bedsections, etc.

DIAGNOSES

Comorbid Psychiatric and Nonpsychiatric Diagnoses and Indicating Diagnoses: Indicator variables were used to reflect a variety of specific psychiatric diagnoses given in the past year, based on ICD-9-CM. We required all cohort members to have VHA service use in the last year as well as a prior year, so this time period maximized information about what diagnoses a patient likely actually had. The one exception was diagnoses that served as an indication for treatment (mood or psychotic diagnoses), for which our criteria was more stringent: we required the diagnosis to be entered in the last 30 days. This was done in order to maximize the likelihood that this was the reason the patient was receiving lithium or valproate.

Nonpsychiatric diagnoses are also of importance to address. In a meta-analysis of literature up through 1993, Harris and Barraclough[64] observed that 19 different nonpsychiatric illnesses were significantly associated with increased suicide risk. Nonpsychiatric diagnoses were aggregated into larger categories based on the comorbid illness categories that make up the Charlson Comorbidity Index and the Elixhauser Comorbidity Index, as per a classification procedure developed for use with administrative databases.[131] For the Charlson index categories, the following 13 (out of the total 17) comorbidity categories were used: Myocardial infarction, Congestive Heart Failure, Peripheral Vascular Disease, Cerebrovascular Disease, Chronic Obstructive Pulmonary Disease, Connective Tissue Disease, Peptic Ulcer, Mild Liver Disease, Moderate or Severe Liver Disease, Diabetes Mellitus without complications, Diabetes Mellitus with complications, Renal Disease, AIDS/HIV Infection.

Elixhauser Comorbidity categories were also included, based on the same reference,[131] when these categories were judged not to overlap with the Charlson index categories. The eleven categories included were: Arrhythmias, Weight Loss, Coagulopathies, Pulmonary Circulation Disease, Hypertension without Complications, Hypertension with Complications, Valvular Disease, Neurodegenerative Diseases, Hypothyroidism, Obesity, Anemia from Blood Loss, and Deficiency Anemia.

Multiple indicators were also included to reflect total score on the Charlson Comorbidity Index, considering all diagnoses received in the past year.

In addition, indicators for other injury-related and a few specific diagnoses that have been linked to suicide risk (progressive, neurodegenerative or autoimmune conditions, and pain diagnosis were included). Finally, an aggregated smoking indicator was included in this category. Tobacco dependence is recognized as being underdiagnosed in VHA administrative/clinical coding, so we constructed a "recent smoking" variable which assumed a value of "1" if a patient had any of the three in the past year: a diagnosis of Tobacco Dependence, at least one visit to a smoking cessation clinic, or prescription of nicotine replacement therapy or varenicline.

Comorbid Substance Abuse Diagnoses: Seven categories of legal/illicit substance use (alcohol, amphetamine, cocaine, marijuana, opioids, sedatives, other substances) were coded as four different indicators reflecting diagnoses received in the last year: dependence on that particular substance, abuse of that particular substance, remission from dependence of that substance and remission from abuse of that substance. The eighth category, hallucinogens, was coded as only 3 indicators (dependence, abuse, and remission from dependence) because there were insufficient numbers of patients (≤ 5 in one of the treatment groups) diagnosed with remission from hallucinogen abuse in the past year. In addition, indicators were included for combined substance dependence and remission from combined substance dependence, including separate indicators denoting whether this combined dependence included opioids or not. Two indicators were also included for "unspecified" substance dependence. Lastly, indicators were included in this category for alcohol intoxication (both a narrow and broad definition) and alcohol or drug psychoses.

Recent Nonfatal Suicidal Behavior Diagnoses: Episodes of nonfatal suicidal behavior, especially those occurring recently, are among the strongest documented risk factors for suicide,[65, 132] however there are concerns that diagnoses may incompletely capture actual episodes of nonfatal suicidal behavior.[133] There are also concerns that outpatient suicidal behavior diagnoses may reflect a history of more remote suicidal behavior rather than behavior necessarily occurring close to the time the diagnoses were entered. To address these concerns a hierarchy was imposed to avoid double-counting of nonfatal suicide behavior episodes between diagnoses recorded during nonmental health hospitalizations, mental health hospitalizations, or during outpatient encounters. Indicator variables were developed reflecting the occurrence of a diagnosis of an episode of nonfatal suicidal behavior over the last 30 days, days 31 to 180 and days 181 to 365 prior to lithium/valproate start. This approach is expected to result in only an approximate indicator of recently diagnosed episodes of suicidal behavior, since a patient could have two separate attempts within a time period that were diagnosed in different settings, and this occurrence would not be reflected in our coding scheme. In addition, the same attempt, if a diagnosis occurred close to the end of a time interval in one setting (e.g., during a non-MH hospitalization), may have been re-diagnosed in a second setting in the next time interval. Thus this single behavior episode would appear as two distinct episodes in our coding scheme, not one. Some imprecision of this type is likely unavoidable.

Despite such uncertainties, given the extreme importance of nonfatal suicidal behavior to predicting suicide risk, we felt it was important to incorporate this information when available in our extensive propensity score. Similarly, it was considered important to maintain this distinction concerning the setting of the nonfatal suicidal behavior diagnosis, since an episode diagnosed in a non-mental health hospitalization is likely to be, on average, considerably more serious than diagnoses simply recorded as outpatient diagnoses. It should be recognized that in general diagnoses of nonfatal suicidal behavior are specific but very insensitive, [133] although this

sensitivity is expected to increase for inpatient diagnoses compared with outpatient (another reason that we made this distinction).

MEDICATIONS

Current and Recent Mental Health Medications: Mental health medication prescriptions active at the time of lithium/valproate start or recently filled (within the last 180 days) were designated into general classes by 24 indicator variables, using a classification system previously developed. This system already uses multiple categories to index antidepressants; for this study we also classified second generation antipsychotics into individual medications (clozapine, olanzapine, risperidone/paliperidone, quetiapine, aripiprazole, ziprasidone). Such an enhanced classification was important given the differential impacts of these medications on both suicide and other mortality risk. An identical number of indicator variables were used to reflect recent but not current prescriptions of medications from these same classes, designating receipt of one or more prescription of that type of medication in the last 180 days in the absence of a prescription whose days' supply includes the start date for lithium/valproate treatment.

For nonmental health medications, a system was developed using medication class code information assigned by the VHA by the VHA national formulary. The VHA assigns every medication administered from the pharmacy into one of more than 1000 classes of medication denoted by the VHA through 5 character "medication class" codes. We took advantage of this classification as a method to logically aggregate prescriptions for related medications (e.g., different thiazide diuretics were able to be aggregated through these codes into a "thiazide diuretic class," different loop diuretics into a "loop diuretic" class, etc.). In many cases, we condensed this "class code" into a 3 character "superclass" code, but in other cases, such as the diuretic example above, in which further distinctions concerning different types of diuretics were judged important, the entire 5 character class code was used. This condensed the approximately 1000 VHA medication classes used by our cohort down to approximately 225 classes/superclasses. Then all revised medication classes present with a prevalence of $\geq 5\%$ in either treatment group (reflecting number of patients with at least one prescription in the last 180 days, or with a current prescription on start date of lithium/valproate) were included, along with any revised medication classes of $< 5\%$ prevalence but $> 1\%$ prevalence that were judged *a priori* particularly relevant to either suicide or other mortality risk (e.g., warfarin, digoxin, etc.).

Indicators for "Current" medication classes required the patient to have an active prescription with days' supply that included the start date of lithium/valproate, while indicators for "Recent" medication classes required the patient to have had at least one prescription filled in the last 180 days but no active supply at time of lithium/valproate start.

In the rare cases when fewer than 5 individuals had received medications of a particular class currently or recently, this class was either removed from the propensity score model or consolidated with other medication classes. This resulted in small

differences, for instance, in the number of classes of current nonmental health medications (54 variables) versus recent medications (55 variables).

Prior Mood Stabilizer Treatment History: Although we sought to identify incident users through the requirement of a "clean period," some patients, although a clear minority (36% of a treatment group or less), had had past treatment more remotely with either mood stabilizers of any type, or specifically with lithium or valproate. Two indicator variables were included, reflecting past history of treatment with mood stabilizers in general and past history of treatment with either lithium or valproate.

OTHER

NonMental Health Diagnoses Possibly reflecting suicide attempts, NonMental Health Utilization of special relevance to suicide risk, and NonMental Health medications of special relevance to suicide risk: Because injuries may occur that are not recognized as representing suicide attempts, we included indicators based on a variety of injury diagnosis codes, reflecting occurrence of these codes in the last year. These indicators included general indicators reflecting any acute injury or any fracture, as well as very specific injuries of concern, such as blood vessel injury, poisoning, and inhalation/drowning/and asphyxiation injury. We also include indicators designating pain clinic use, opiate pain medication use, and designating if patients had received activated charcoal, or naloxone or flumazenil in the past year.

Geographic Suicide Risk: Indicator variables were constructed to classify patients into 5 categories (approximate quintiles) of age-adjusted regional (state-level) suicide risk, based on publically available data from the Centers of Disease Control, which was available from 2000-2007.[134] Because these statistics would include the suicides of Veterans occurring in this period, there is a theoretical potential for some bias to be introduced by control of this covariate. However, practically, this bias is expected to be exceedingly small, given that >150,000 suicides occurred across these states over eight years, and our sample accounted for only 102 suicides over that period (< 0.1%). A geographic suicide risk indicator was included because suicide risk has been found to vary substantially from state to state for reasons that are not completely understood but that might be also expected to influence suicide risk in Veterans specifically (e.g., access to firearms).

NonMental Health Diagnostic Testing: Clinic stop codes reflecting diagnostic procedures over the last 180 days and days 181 to 365 prior to lithium/valproate start were used to construct indicators of the frequency of diagnostic tests over the past year: X-Rays, CT or MRI scans, EKGs, Ultrasound, Echocardiograms, Endoscopy, Pulmonary Function Tests (PFTs), Nuclear Medicine, and Angiograms (for Angiograms, tests were divided as occurring within the last 180d days and in days 181 to 365 prior to lithium/valproate start).

Three additional variables were included to help balance the extensiveness of pharmacy records among our recipients: any prior use of VA pharmacy, use > 180 days prior to LI/VAL start, and use > 365 days prior to LI/VAL start.

The Table following this Appendix (Appendix 1-2 Supplementary Table 1) illustrates how the extensive propensity score-matching strategy balanced the treatment groups on key measured covariates. Because of the much greater number of valproate recipients in our unmatched cohort, the effect of the matching is essentially to select those valproate recipients most similar (in measured covariates) to the lithium recipients. For instance, the single covariate most imbalanced between treatment groups in the unmatched cohort (Bipolar I diagnosis, with a standardized difference of 0.28 between treatment groups) is much more closely balanced in the matched sample, with the two groups having a highly similar prevalence of Bipolar I diagnosis (45.1% versus 45.7% for a standardized difference of 0.011) that are close to the prevalence of Bipolar I diagnosis in the original, unmatched sample of lithium recipients. In this fashion, the extensive propensity score matching produced a sample from within the original unmatched cohort closely balanced (all standardized differences after matching < 0.018) on all 934 covariates.*

---

* Note concerning covariate count: In the manuscript and here, we refer to 934 covariates because these were the number of separate, unique quantities balanced through the extensive propensity score matching. This includes "0 count" indicators for the variables modeled as more than 2 levels (i.e. more than just absent/present). For variables with > 2 levels, but not dichotomous variables, the number of individuals lacking any presence of that indicator (e.g., 0 additional psychiatric medications at baseline) is a separate quantity, rather than simply another form of the information that can be obtained from the count of individuals scoring "1" for the indicator.

Appendix 1-2 Supplementary Table 1. Key Characteristics of Patients Initiating Lithium (Li) and Valproate (VAL) both Prior to and After Propensity-Score Matching[a]

	UNMATCHED Sample			MATCHED Sample		
Characteristic	**Li (n=21468)** n, (%)	**VAL (n=71887)** n, (%)	**Std. Diff.[b]**	**Li (n=21194)** n, (%)	**VAL (n=21194)** n, (%)	**Std. Diff.[b]**
Demographics						
Age 50+[c]	10353 (48.2)	36435 (50.7)	0.049	10244 (48.3)	10156 (47.9)	0.008
Sex (Female)[d]	2978 (13.9)	6750 (9.4)	0.140	2894 (13.7)	2934 (13.8)	0.005
Race, White	16994 (79.2)	52493 (73.0)	0.144	16748 (79.0)	16793 (79.2)	0.005
Race, Black	2833 (13.2)	14197 (19.7)	0.177	2825 (13.3)	2770 (13.1)	0.008
Married	7500 (34.9)	26484 (36.8)	0.040	7416 (35.0)	7298 (34.4)	0.012
State Suicide Rate, 3rd quintile	3325 (15.5)	14647 (20.4)	0.128	3305 (15.6)	3251 (15.3)	0.007
Indicating Diagnosis[e] (Past 30 days)						
Bipolar I	9737 (45.4)	22811 (31.7)	0.283	9562 (45.1)	9683 (45.7)	0.011
Bipolar NOS	1686 (7.9)	3630 (5.0)	0.114	1643 (7.8)	1661 (7.8)	0.003
Depression NOS	4233 (19.7)	21693 (30.2)	0.243	4214 (19.9)	4129 (19.5)	0.010
Schizophrenia	924 (4.3)	6605 (9.2)	0.196	924 (4.4)	949 (4.5)	0.006
Other Psychosis	252 (1.2)	1914 (2.7)	0.109	252 (1.2)	255 (1.2)	0.001
Additional Psychiatric Diagnoses (Past Year)						
PTSD	4894 (22.8)	20011 (27.8)	0.116	4842 (22.8)	4749 (22.4)	0.010
Alcohol Dep	4499 (21.0)	15713 (21.9)	0.022	4426 (20.9)	4478 (21.1)	0.006
Suicidal Behavior Diagnoses (Suicide Attempt) (past 30d, by location where diagnosed (Dx))						
NonMH Hosp Dx	28 (0.13)	122 (0.17)	0.010	28 (0.13)	24 (0.11)	0.005
MH Hosp Dx	30 (0.14)	129 (0.18)	0.010	30 (0.14)	32 (0.15)	0.002
Outpatient Dx	145 (0.68)	507 (0.71)	0.004	144 (0.68)	147 (0.69)	0.002

Appendix 1-2 Supplementary Table 1. (continued)

Suicidal Behavior Diagnoses (Suicide Attempt) (past 31-180d)						
NonMH Hosp Dx	44 (0.20)	89 (0.12)	0.020	43 (0.20)	43 (0.20)	0.000
MH Hosp Dx	32 (0.15)	87 (0.12)	0.008	31 (0.15)	29 (0.14)	0.003
Outpatient Dx	91 (0.42)	276 (0.38)	0.006	90 (0.42)	82 (0.39)	0.006
Possible Suicidal Behavior-Related Diagnoses (past year)						
Any Acute Injury	3950 (18.4)	13569 (18.9)	0.012	3872 (18.3)	3884 (18.3)	0.001
Psychiatric Hospitalizations						
D/C past 7 days	2260 (10.5)	9821 (13.7)	0.096	2232 (10.5)	2219 (10.5)	0.002
D/C past 8-30d	879 (4.1)	3469 (4.8)	0.035	863 (4.1)	881 (4.2)	0.004
D/C Past 31-180d	2062 (9.6)	7293(10.1)	0.018	2024 (9.5)	2063 (9.7)	0.006
Current Psychiatric Medications						
Other Mood Stabilizer(s)	3009 (14.0)	6875 (9.6)	0.138	2891 (13.6)	2854 (13.5)	0.005
SSRI antidep	7700 (35.9)	28496 (39.6)	0.078	7615 (35.9)	7666 (36.2)	0.005
SNRI antidep	2046 (9.5)	4993 (6.9)	0.094	1988 (9.4)	2019 (9.5)	0.005
Past Treatment History						
Prior Mood Stabilizer	7680 (35.8)	20795 (28.9)	0.147	7503 (35.4)	7530 (35.5)	0.003
Diagnoses, Nonpsychiatric (past year)						
Mild Liver Dz	1892 (8.8)	3308 (4.6)	0.169	1747 (8.2)	1719 (8.1)	0.005
Outpatient Utilization, Nonpsychiatric (past 180d)						
Gastroenterology Clinic, 1+ visits	1197 (5.6)	2466 (3.4)	0.104	1102 (5.2)	1077 (5.1)	0.005

Appendix 1-2 Supplementary Table 1 (continued)

Current Medications, Nonpsychiatric						
Thiazide Diuretic	1515 (7.1)	7650 (10.6)	0.126	1499 (7.1)	1492 (7.0)	0.001
ACE Inhibitor	2784 (13.0)	12320 (17.1)	0.117	2764 (13.0)	2736 (12.9)	0.004
NSAIDs	3516 (16.4)	14738 (20.5)	0.106	3491(16.5)	3522 (16.6)	0.004

[a] A partial version of this Table appears as Manuscript Table 2. Since the degree of imbalance in these variables occurring prior to matching may be of interest to some readers, we present this Table again with 5 extra columns to report the prevalence of these covariates in the sample prior to matching, and to show the reduction in imbalance resulting after the extensive propensity score matching.

[b] Std. Diff. = Standardized Difference.

[c] Age presented in this format (<50 years old vs. ≥50 years old) to streamline its presentation within this Table: age was actually modeled using 11 indicators reflecting age groups from <35 years old in 5-year intervals to ≥ 80 years old.

[d] The proportion of females in the cohort is low because the veteran sample is predominantly male.

[e] Percentages for Indicating Diagnoses do not add up to 100% because some diagnoses are not substantially imbalanced and therefore not listed in this Table (e.g., Major Depression, Bipolar II Disorder, ≥2 Indicating Diagnoses in past 30 days), although they were included in the propensity score and balanced through matching.

ABBREVIATIONS: Dep = Dependence; D/C =Discharge, NonMH Hosp Dx = Diagnosed during a Non-Mental Health hospital stay, MH Hosp Dx = Diagnosed during a Non-Mental Health hospital stay, Outpatient Dx = Diagnosed during an outpatient visit, SSRI = Serotonin-Specific Reuptake Inhibitor, antidep = antidepressant, SNRI = Serotonin-Norepinephrine Reuptake Inhibitor, Dz = Disease.

Appendix 1-3. Mental Health Medication Covariates Included in the Analysis

Because psychiatric medications are of particular importance in both helping to index the severity of various psychiatric diagnoses and also as potential direct influences on suicidal behavior (e.g., clozapine), we sought to control for a wide variety of potential psychiatric medications that cohort members might be receiving. We also sought to produce, through propensity score-matching, two cohorts that were not only similar in the psychiatric medications that patients were currently receiving, but also medications that they have recently been receiving (within the last 6 months) but were not receiving currently. Such medications may have been treatments that they or their provider deliberately decided to stop, or intended to continue but were not successful in so doing, or for which they were experiencing only a brief interruption in treatment that happened to occur in proximity to their lithium/valproate treatment initiation date. An additional reason it is important to control for concomitant medications is that a current or recent history of receiving a psychiatric medication may also influence the subsequent psychiatric medications a patient might receive.

Table 1 in the manuscript lists only those medications for which a substantial initial imbalance occurred between the treatment groups. Appendix 1-3 Supplementary Table 1 below lists all the psychiatric medications or medication classes that were controlled in our analysis; each of these categories was balanced between the treatment groups to a standardized difference of < 0.018 for each of the time periods.

Appendix 1-3 Supplementary Table 1. Listing of Mental Health Medications that were Propensity Score-Matched between the Lithium and Valproate Treatment Groups
(24 medication/medication classes x 2 time periods)

CURRENT MENTAL HEALTH MEDICATIONS *(24 variables)*
(active prescription on LI/VAL start date)

Other Mood Stabilizers (carbamazepine, lamotrigine, etc.)	TCAs
Olanzapine	MAOIs
Quetiapine	Benzodiazepines
Risperidone	Other Hypnotics
Ziprasidone	Buspirone
Aripiprazole	Stimulants
Clozapine	Disulfarim (Antabuse)/Naltrexone
First Generation Antipsychotics	Buprenorphine
SSRIs	Methadone
SNRIs	Antihistamines
Bupropion	Anticholinergics
Mirtazapine	Atypical Dopaminergic medications

RECENT MENTAL HEALTH MEDICATIONS *(24 variables)*
(active prescription within the last 180 days but no prescribed supply extending to LI/VAL start date)

Same medications/medication classes as Current Mental Health Medications

Appendix Table 1-1. Rates of Continuation or Discontinuation of Initial Treatment and Other Censoring Events, by Treatment[a]

Treatment Status	90-day Follow-up				180-day Follow-up				365-day Follow-up			
	Lithium		Valproate		Lithium		Valproate		Lithium		Valproate	
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Still Receiving Initial Treatment	9920	(46.8)	9950	(47.0)	4987	(23.5)	5145	(24.3)	1612	(7.6)	1712	(8.1)
Discontinued Initial Treatment	10455	(49.3)	10858	(51.2)	15146	(71.4)	15532	(73.3)	18344	(86.6)	18873	(89.1)
Initiated opposite mood stabilizer[b]	772	(3.6)	327	(1.5)	992	(4.7)	426	(2.0)	1150	(5.4)	489	(2.3)
Died from Other causes	32	(0.15)	42	(0.20)	52	(0.25)	74	(0.35)	70	(0.33)	99	(0.47)
Died from Suicide	15	(0.07)	17	(0.08)	17	(0.08)	17	(0.08)	18	(0.08)	21	(0.1)

[a] n= 21194 propensity-score matched pairs.

[b] Patients may have reinitiated treatment subsequently with the same or different mood stabilizer, but this occurred after being censored from the “Still Receiving Initial Treatment” subsample of our Intent-to-Treat cohort due to a gap in treatment.

[c] This count provides the number censored due to an immediate switch to the other mood stabilizer.

.

Appendix 1-4. Survival Analysis of Suicide Risk by Treatment over 0-90 days, 91-180 days, and 181-365 days

Survival analysis was performed on the intent-to-treat sample using standard Cox regression techniques, but its interpretation was complicated by the fact that "nonproportional hazards" over the 0-365 day time period were observed. The observation of "nonproportional hazards," as evidenced by the crossing of the survival curves at approximately 90 days (Manuscript Figure 1) and a statistically significant time*treatment interaction term ($p = 0.03$) means that an important assumption of the Cox model was not met. As a response to the observation of nonproportional hazards, we adopted one of several established approaches to addressing nonproportional hazards: segregating follow-up time into periods over which proportional hazards were observed.[135]

However, this segregation of time has an important ramification. Some of the follow-up periods have initiation dates after the actual treatment initiation date, which means the close balancing of the treatment groups in the propensity score covariates at day 0 can no longer be presumed to necessarily hold for later time periods. This is a substantial limitation. However, it should be noted that for readers interested in examining the differences in suicide risk associated with lithium and valproate treatment that accounts for differences in amounts of follow-up time (usually a key function of survival analysis), such results are, for practical purposes, already provided in Tables 3 and 4 of the manuscript. This is because the extensively propensity-score matched treatment groups exhibited highly similar rates of treatment discontinuation for the 0-365 day and briefer time period analyses. For instance, for the primary (intent-to-treat analysis) over 0-365 follow-up time was 7,699, 086 person-days for the lithium treatment group and 7,690,014 person-days for the valproate treatment group (a difference of 0.1%). Thus, analyses that focus on events and patients (logistic regression) provide very similar results to analyses which formally incorporate person-days of exposure, as the very close agreement between the logistic regression and rate ratio results provided in Manuscript Tables 3 and 4 demonstrates.

Given these considerations, we present the logistic regression in the manuscript and the survival analysis results here (Appendix 1-4 Supplementary Table 1). The survival analysis segregated follow-up time into periods 0-90 days, 91-180 days, and 181-365 days after medication initiation to account for nonproportional hazards. Significantly increased risks of suicide death were observed among all patients initiating lithium, but only for the 91-180 day time period (Hazard Ratio (HR) = 3.50, 95% CI 1.41, 8.66; Appendix 1-4 Supplementary Table 1). Similar to the 0-180 day logistic regression results reported in the manuscript, virtually all elevated suicide risk among patients initiated on lithium during this time period occurred among patients who had stopped or modified lithium treatment (19 out of 21 suicides, HR = 3.14, 95% CI 1.25, 7.85). Risks among patients stopping/modifying lithium treatment were less pronounced over other time periods, especially after completion of the first 180 days of follow-up (181-365 HR (after stopping/modifying treatment) = 0.93, 95% CI 0.47, 1.80). This lack

of increased risks among patients stopping/ modifying treatment over 181-365 days occurred in conjunction with distinctly, although nonsignificantly, reduced risks among patients still receiving initial lithium treatment (HR = 0.26, 95% CI 0.03, 2.34, p = 0.23). Although based on extremely few suicides (1 versus 4), these results for active treatment with lithium after 180 days may be consistent with the suggestion that lithium will reduce suicide over longer durations. Such an association could be becoming apparent in our analyses over this period either due to an increasing protective effect of lithium over time or due to diminishment of time-varying confounding in our analyses biasing against lithium.

Given the unusual size of our cohort, to benefit future research we explored whether the nonsignificant association of active lithium treatment after 180 days strengthened if we lengthened follow-up further. Interestingly, the association of lithium with reduced suicide risk among patients still receiving initial treatment strengthened and reached borderline statistical significance from 181–730 days (1 suicide among lithium-treated and 6 suicides among valproate-treated patients, HR = 0.18, 95% CI 0.02, 1.45, p = 0.11), but not from 181 days until the end of the study period (up to 10+ years for some patients) (8 suicides among lithium-treated and 8 suicides among valproate-treated patients, HR = 1.09, 95% CI 0.41, 2.89). As follow-up time progressively lengthens, the lack of reweighting over time would be expected to be an increasingly strong limitation (that is, the small fraction of patients still receiving lithium after several years of treatment might be considerably different than the fraction of patients continuing to receive valproate). Clearly, still larger cohorts or cohorts with substantially greater rates of treatment persistence will be needed to reliably examine the associations between lithium and valproate treatment and suicide risks over follow-up times longer than 365 days. In addition, alternative designs could be considered to facilitate examinations of longer follow-up times, such as examining the more selectively reported but numerous outcome of nonfatal suicidal behavior and using marginal structural models to reweight patient samples during follow-up.

Appendix 1-4 Supplementary Table 1. Cox Regression Survival Analysis by Time Period since Medication Initiation

Intent-to-Treat Cohort	
Time Period	Hazard Ratio (95% CI)
0-90 days	0.95 (0.60-1.50)[a]
91-180 days	3.50 (1.41-8.66)[b]
181-365 days	0.81 (0.43-1.53)

Stratified by Treatment Status		
Time Period	Hazard Ratio (95% CI)	
	During Exposure to Initial Treatment	After Stopping/Modifying Initial Treatment
0-90 days	0.93 (0.54-1.58)	1.43 (0.24-8.36)
91-180 days	NC[c]	3.14 (1.25 – 7.85)[d]
181-365 days	0.26 (0.03-2.35)[e]	0.93 (0.47-1.84)[f]

[a] Based on Stratified Cox regression (stratified on matched pairs), all other Hazard Ratios non-stratified.

[b] p = 0.007. Other Intent-to-Treat comparisons (0-90 days and 181-365 days were not significant at 0.05 level).

[c] NC = "Not calculable." A hazard ratio cannot be calculated because of 0 suicides in the valproate subcohort still receiving initial treatment over this period (2 in lithium subcohort still receiving initial treatment).

[d] p = 0.015. Other comparisons during exposure to initial treatment and after stopping/modifying initial treatment not significant at the 0.05 level.

[e] Based on 5 suicides (1 in lithium and 4 in valproate subcohorts still receiving initial treatment).

[f] Based on 33 suicides (16 in lithium and 17 in valproate subcohorts discontinuing/modifying treatment).

Appendix 1-5. A Potential Integration of Key Study Findings

Because this study is distinct in its size compared to past studies, it is worthwhile to extract as much information from this study as feasible that may helpfully inform judgments concerning its clinical and research implications. The sections below are intended to provide a succinct summary of the reasons why the study results appear, in a strict sense, to be compatible with four different scenarios: 1) general equivalency between the treatments, 2) increased suicide risks associated with lithium discontinuation, 3) decreased suicide risks associated with active lithium treatment, 4) and, in what we suspect to be the most likely scenario, a combination of some degree of decreased risks during active treatment and some degree of increased risks upon lithium discontinuation. Such a scenario would also appear to be consistent with substantial past literature, given that many nonrandomized studies reporting associations between lithium treatment and reduced risks of suicide were restricted to patients receiving active treatment, and several studies have documented dramatically increased risk of suicide or suicidal behavior shortly after the discontinuation of lithium.[13, 20, 21, 23, 136] Nevertheless, it is not clear how much more likely this interpretation is to be true than an interpretation that posits little or no change upon active treatment but distinct risks upon discontinuation, or an interpretation that posits substantial decreases in risk during active treatment and generally equivalent risks upon discontinuation. The final possibility, equivalency between the treatments, both during active treatment and after discontinuation appears the least likely, given the significant risks observed over the 0-180 and 91-180 days periods virtually exclusively associated with lithium discontinuation, and the significant differences in diagnostically-coded suicidal ideation.

Nevertheless, because this synthesis requires integration of both statistically significant, borderline significant, and (at some points) clearly nonsignificant results, the level of confidence to be placed in this interpretation is highly uncertain. This material should be viewed as a qualitative and nondefinitive synthesis of the overall study findings. It is intended to inform interpretations about the most likely clinical and research implications of this study, without attempting to quantitatively estimate the degree to which these interpretations are more likely than the specific alternatives that are discussed.

Appendix 1-5A. Implications of Observed High Treatment Discontinuation Rates

The rates of treatment discontinuation observed in the study cohort are quite substantial. In general, the rates of treatment impersistence in this study appear to equal or exceed those reported previously. However, many of these reports do not investigate comprehensive incident cohort samples.[97, 109, 110, 137] Three exceptions are the Johnson and McFarland study[113] examining all patients initiating lithium in an HMO, the Kessing et al. study[114] examining all patients initiating lithium in Denmark, and the Licht et al. lithium clinic study.[138] Johnson & McFarland[113] found discontinuation rates slightly greater than our study (median time to discontinuation 72 days, rather than approximately

90 days in our cohort), whereas Kessing et al.[114] found rates slightly lower rates (82% discontinuation in one year rather than the approximately 92% discontinuation in a year observed in our study). Both these studies used estimates of prescription length based on number of pills, rather than using the actual prescription directions to calculate a days' supply as done in this study. In addition, Kessing et al.[114] mentioned their sample might have had lower than average illness severity given that a majority of prescriptions were provided by general practitioners, not psychiatrists. Also, treatment discontinuation rates may be higher in our Veteran sample than other samples, given the high rates of comorbidities, substance use disorders [139], and homelessness. Licht et al.[138] reported only a 19% discontinuation rate for lithium treatment over 2 years; however, their sample was from a specialized lithium clinic and their definition of discontinuation was not provided.

Despite the fact that quite substantial rates of treatment discontinuation are observed, the observed rates are highly similar between the two treatments. For example, at 90 days 46.8% of patients initiating lithium and 47.0% of patients initiating valproate remained on initial treatment, at 180 days 23.5% versus 24.3% patients respectively, and at 365 days 7.6% versus 8.1% of patients, respectively. The fact that rates of discontinuation were highly similar between the two treatment arms is reassuring in one important sense. Given the initial close balance in measured factors, if substantially different treatment discontinuation rates between the treatment groups had been observed instead, this would immediately suggest that the patients remaining on initial treatment in each treatment group differed more substantially on measured suicide risk factors than at initiation. However, while the similarity in rates is reassuring, confirmation that the *reasons* for discontinuation are similar is necessary to firmly conclude that treatment discontinuation occurring during follow-up did not substantially affect the covariate balance between the treatment groups.[48] Such information is often not available,[48] especially if selection to discontinue may relate to poorly measured or unmeasured factors (e.g., information about suicidal planning, symptoms such as hopelessness, etc.).

Nevertheless, concomitant psychiatric medications are one class of measured factors which could potentially exert some degree of influence on this study's findings. Given that this study's objective was to characterize whether a comparative difference existed in the suicide risk associated with lithium and valproate in their associations with suicide risk among Veterans receiving usual VHA treatment, the study cohort was not restricted to individuals receiving strict monotherapy. Differences between the treatment groups in concomitant psychiatric medications could arise after lithium or valproate initiation, for instance if a tendency existed for patients in one group to be more likely to add medications during treatment or after initial treatment discontinuation. Because other psychiatric medications may have their own relationship to suicide risk, rigorous examination of these possibilities would ultimately be desirable, and we recommend this be a focus of future research. However, approaches such as marginal structural models which periodically reweight samples based on measured factors to keep the treatment groups closely comparable may potentially be susceptible to the same amplification of the effects of unmeasured factors as baseline confounding; therefore, explorations of the effects of concomitant psychiatric medications may need to combine or contrast findings

from marginal structural models with approaches that restrict samples to individuals not receiving any concomitant psychiatric medications, or certain psychiatric medications, at any point during follow-up. Because such restrictions reduce sample size, such efforts ideally would involve even larger samples than the one we had available.

It should be recognized that close balancing was achieved of the patient groups on an extensive set of medications psychiatric present at the time of lithium or valproate initiation (Appendix 1-3), and also on wide variety of other factors. Thus, at least for a substantial portion of the follow-up period (the earlier months), concomitant medications between the groups are likely to be highly similar. In addition, the balance achieved between the treatment groups also included a number of fixed factors (e.g., age) and slowly time-varying factors (e.g., additional psychiatric diagnoses) that might influence prescribing of concomitant medications. Furthermore, a distinct aspect of this study was that not only were current concomitant psychiatric medications controlled, but also the use of psychiatric medications within the last 6 months that were no longer being currently prescribed. To the extent that recent receipt of particular classes of medications might plausibly influence choice concerning what subsequent medications should be initiated, these influences were tightly balanced at baseline. As mentioned above, a design that rebalances the sample during follow-up with respect to concomitant medications is the ideal and should certainly be a future research priority, especially if efforts are made to examine longer periods of follow-up. This study did incorporate design features which likely limited the impact of selection during follow-up compared to some other designs. However, since we did not formally attempt to control the impacts of selection during follow-up, this limitation should be borne in mind when interpreting this study.

While our study did not attempt a formal analysis incorporating changes occurring during follow-up, some aspects of our study serve to diminish concerns that differences between the treatment groups in the selection of patients to continue or discontinue treatment during follow-up explain all or most of our findings. Specifically, differences in selection during treatment alone would not be expected to create the (marginally significant) differences in intent-to-treat estimates measured from treatment initiation. Intent-to-treat estimates continue to incorporate outcomes from all treatment initiators, regardless of a patient's status of still receiving or having discontinued initial treatment. In a sense, patients are not selected out of an intent-to-treat cohort during follow-up. This is especially for an outcome such as suicide, which is comprehensively documented nationwide, regardless of whether patients continue to receive care from the Veterans Health Administration.

Differences in selection during follow-up would certainly be expected bias the treatment effect estimates observed for patients during active treatment (the only type of treatment effect estimate reported in virtually all of the earlier nonrandomized studies of lithium) or after treatment discontinuation. However, such selection would not bias the intent-to-treat estimates (in the absence of genuine treatment effects). Thus, the marginally statistically significant intent-to-treat results observed associated with lithium treatment after discontinuation over 0-180 days, and the significant risks result over 91-

180 days (Appendix 1-4), are very important observations since they help restrict the potential explanations for the significantly elevated risk observed in patients discontinuing lithium over 0-180 days.

The high rates of discontinuation also suggest the possibility that discontinuation decisions may have been made largely by patients rather than providers, and these decision may have plausibly related much more to concerns such as stigma, side effects, and a lack of perceived efficacy than suicide risk directly. Thus, it is possible that the high rates of discontinuation, while quite sizable, may have had relatively little impact on the treatment effect estimates. In addition, any early provider-based selection that occurred should probably most likely be suspected of occurring in the direction of initial confounding, rather than against it. Initial confounding appears to bias to some degree towards higher risks observed with lithium, based on the risks observed in the unmatched cohort and the greater prevalence of diagnostically-coded suicidal ideation among patients initiating lithium in the matched cohort. If providers on average selected higher risk patients to initiate lithium at Day 0, it seems somewhat unlikely that they would reverse this tendency in treatment very shortly after treatment started. Selection in the direction of more often retaining higher-risk individuals on lithium than valproate would be in the opposite direction of what would be needed to explain the significantly elevated risks in patients discontinuing lithium over 0-180 days.

As a side note, if selection during follow-up did occur in the direction of retaining higher-risk individuals on lithium, this would represent a likely *third* process, along with residual baseline confounding and the likelihood, discussed in the manuscript that some suicides attributed as occurring "during initial treatment" actually occurred after treatment discontinuation, that would be expected to lead to an underestimate of the benefit of lithium during active treatment.

Appendix 1-5B. Likelihood of Substantial Residual Confounding Amplification

Propensity score designs, especially when studying rare or infrequent outcomes, permit inclusion of far more covariates than some alternative approaches. We sought to take advantage of this capability to thoroughly control for numerous suicide risk factors and potential suicide risk factors in this study's design by including a large variety of covariates and flexibly modeling their distribution, frequency and timing. We recognized that several potential suicide risk factors (suicidal ideation, planning, etc.) would remain unmeasured. If unmeasured confounding remains uncontrolled in a propensity score analysis, it has recently become appreciated that the effect estimates produced may include an additional source of balance: amplification of whatever confounding remains uncontrolled after application of the propensity score methods.[43-46] This problematic effect occurs if the propensity score includes covariates with a substantial or strong association with treatment exposure in the absence of an association with outcome. We took steps that limit the amount of potential amplification of residual confounding that application of our propensity score approach might produce. All variables were evaluated to determine their relationship with both treatment and outcome. Covariates

with particularly substantial relationships with exposure to one or the other treatment were then evaluated individually to assess their plausibility as confounders. (Of note, none of the many covariates in the model had what might be considered a particularly "strong" association with treatment exposure by some definitions. That is, no covariates had an odds ratio for treatment exposure to lithium versus valproate, of even 3.0 or 0.33). Nevertheless, given the large number of covariates included in the model, it is possible that some degree of amplification of residual confounding was produced by our design.

However, even if the design created the potential for some amplification of residual confounding, the actual quantitative bias that would result would depend heavily on how much residual confounding was present after application of the propensity score. If little or no residual confounding exists, amplification of this confounding would have to exist on a very pronounced scale (e.g., 2-fold, 3-fold, etc.) to substantially bias the overall findings (assuming a reasonably-sized treatment effect estimate exists). Importantly, with a c statistic of just 0.69, our propensity score is in the lower portion of the range of exposure prediction. A recent simulation, although using R^2 rather than the c statistics, found that propensity scores in the lower portion of the range of exposure prediction should be expected to amplify confounding somewhat modestly (i.e., < 2-fold).[47] In addition, in our study it remains possible that even the initial confounding may have been fairly minimal, given the generally close balance (standardized difference < 0.1) observed initially for over 98% of the covariates examined.

Nevertheless, we did observe a significant difference in the non-matched covariate denoting the presence of diagnostically-coded suicidal ideation, although the imbalance between treatment groups was only OR = 1.30. Due to the low prevalence of diagnostically-coded suicidal ideation, this corresponds to a standardized difference of only 0.04, which is still greater than any standardized difference for any variable included in the propensity score (< 0.018). Very little is known about the degree to which diagnostic codes for suicidal ideation underestimates actual suicidal ideation, but underestimation almost certainly occurs. However, sensitivity analyses proportionally boosting the prevalence of suicidal ideation indicates that only when the rates for diagnosed-coded suicidal ideation are multiplied ≥ 6-fold to reflect possible overall suicidal ideation rates (i.e., including ideation that is both coded and which was not recorded with the diagnostic code) does a standardized difference of ≥ 0.10 occur. Nevertheless, this would correspond with suicidal ideation rates in the past 30 days of 15-19%. Such a rate conceivably is plausible, albeit perhaps at the upper limit of what might be expected. This sensitivity exercise also presumes that the same difference in suicidal ideation occurs between the treatment groups for the non-coded suicidal ideation rates as for the diagnostically-coded rates.

The possibility that standardized differences even in an important covariate not included in the propensity score may remain generally modest (e.g., standardized difference < 0.10) is important. As pointed out above, the influence of confounding amplification on the results is proportional to the amount of residual confounding that remains. If residual confounding is modest, the added effect of residual confounding amplification is likely to be still more modest, at least in this range of exposure

prediction.[47] However, just as clearly, if a substantial degree of residual confounding persists (in this study such confounding might result from the known suicide risk factors not able to be incorporated in the model), then amplification of residual confounding could potentially increase overall residual confounding. In a sense, amplification of residual confounding would be expected to operate similar to amplification in other systems: if residual confounding is minimal, most levels of amplification will not produce a level of confounding that is much different quantitatively. However, if residual confounding is substantial, then amplification of residual confounding can serve to substantially further increase the degree to which the effect estimate reflects confounding.

Perhaps this study's most important observation regarding the degree of concern that should exist regarding possible amplification of residual confounding amplification is the observation that regardless of possible amplification of residual confounding, the propensity score matching methodology appears to have effectively reduced the *overall confounding* observed between the treatment groups. At each time point studied, the odds ratios obtained prior to matching were further from the null than after matching. For instance, over 0-90 days, when the highest proportion of patients were receiving active treatment, matching on measured factors reduced the central estimate of the intent-to-treat association from 1.10 to 0.95. Over 0-180 days, movement in the intent-to-treat odds ratio estimate from 1.70 to 1.56 was observed after the propensity score matching. Over 0-365 days, the analysis which was informed by the largest number of outcomes, the intent-to-treat odds ratio prior to matching had a central estimate of 1.45, while after matching a central estimate of 1.22 was obtained. Given that the number of past findings suggesting that active lithium treatment is associated with either a reduction or at least a neutral association with suicide risk,[12, 15] such movement in the estimate away from more extreme increased risks being associated with lithium treatment suggests that overall confounding has been reduced, not amplified.

The observation that overall confounding appears to have been reduced by the propensity score matching methodology does not mean confounding amplification resulting from our methodology does not exist, nor that no confounding exists. Rather, this data suggests that any confounding amplification introduced by our propensity score matching methodology, when added to the remaining confounding already present, is not sufficient to negate the effectiveness of the methodology in improving our reported results by beneficially reducing overall confounding. Of course, if substantial residual confounding amplification is present, this implies that a more optimal control of overall confounding is possible. However, the path to achieving that more optimal state is not necessarily obvious. Removal of variables from the propensity score would certainly be expected to reduce residual confounding amplification, but ironically may increase the amount of residual confounding (if the removed variables actually had a recognized or unrecognized association with outcome), so that overall confounding might actually increase.

The risks of suicide by treatment from the unmatched cohort are important to examine for a second reason. They indicate that the general pattern of intent-to-treat risk

observed in the matched analysis closely parallels the pattern observed prior to the propensity score matching. That is, the pattern of generally similar intent-to-treat risks over 0- 90 days, changing to substantially increased risks with lithium treatment at 0-180 days due to a prominence of risks among patients discontinuing lithium, followed by a lessening of this increased risk at 0-365 days, is not a product of some artefact produced by the propensity score matching. It is a pattern observed even prior to any matching, and thus does not appear to result from the actions of any residual confounding amplification.

Another important observation relevant to judgments about confounding amplification results from the sensitivity analysis in which approximately half of the propensity score variables were removed (Appendix 1-6). This modification resulted in only a modest change to the effect estimate. Given that removal of these variables were associated with only a modest change in the treatment effect estimate, the corollary is inclusion of these covariates, despite the fact they exhibited only minimal univariate association with outcome, likely produced only modest confounding amplification. Stated another way, the observation that large risks that continue to be associated with lithium treatment discontinuation despite removal of all covariates lacking a substantial association with outcome (+/- 20%) suggests that either risks associated with lithium discontinuation, nonamplified residual confounding, or possibly selection during treatment is largely responsible for those significantly increased risks, rather than amplification of residual confounding. As we discuss below, this finding does not necessarily mean that overall residual confounding was modest, since we were unable to control for some important risk factors, only that amplification of any residual confounding amplification appears to be modest in effect.

In the future, there are certainly alternative approaches which can be considered when employing an extensive propensity score in a study of suicide risk to potentially optimize confounding control while further limiting confounding amplification. One approach would be to apply an outcome-based selection criteria from the beginning of the study (e.g., such as requiring included covariates have at least a +/-20% association with suicide). In some cases, the lack of a univariate association with outcome does not necessarily indicate that variable is not a genuine confounder. Associations with correlated variables with differing associations with suicide risk could conceal the actual relationship between the variable and the outcome. Alternative approaches might be to adopt the 20% restriction for variables judged particularly unlikely to be associated with suicide risk (e.g., the nonmental health covariates with the least established association with suicide risk), or apply the 20% restriction just to those covariates with the strongest association with treatment, or select the variables on the basis of highly multivariate regression associations. However, the approach which thus far has been demonstrated to apparently minimize confounding thus far in two patient cohorts is a blanket requirement that all covariates have at least a +/- 20% association with outcome,[102] although the generalizability of this observation is uncertain. Another decision point to be explored is how to handle multilevel variables. We retained multilevel covariates in which any strata had at least a 20% association with suicide, but alternatives can be readily envisioned of

requiring that a majority of strata have at least a 20% association, or all strata have such an association.

Clearly, further research in this area is of particular importance. In sum, however, it does not appear that amplification of residual confounding was likely a major influence upon our findings. This tentative conclusion is suggested by the observations that our overall propensity score approach appeared to result in a substantial reduction in confounding, did not alter the basic pattern of risks over time and by treatment status observed between patients initiating lithium and valproate, and the observation that the removal of almost half of our propensity score covariates had only a modest effect on the treatment effect estimates.

Appendix 1-5C. Likelihood of Some Residual Confounding Persisting in the Analysis

Although any *amplification* of residual confounding may be modest, at least three lines of evidence that suggests that some degree of residual confounding may persist in the analysis. The first and simplest line of evidence is that rates of suicidal ideation (as reflected by diagnostic codes received by member of the cohort from 2005-2008) were statistically different between the treatment groups. The difference in prevalence in diagnostically-coded suicidal ideation is modest (OR = 1.30, 95% CI 1.09, 1.54), and patients who express suicidal ideation are not necessarily those at the highest risk of suicide.[140] However, such patients are almost certainly at higher risk for suicide than many other patients in the cohort, thus this data strongly suggests the presence of at least some degree of baseline confounding biasing against lithium. While it is easy to appreciate how the imbalance between the treatment groups in diagnostically-coded suicidal ideation could potentially lead to residual confounding biasing against lithium, estimating the potential quantitative size of this effect is much more difficult. Nevertheless, such estimates, even if somewhat qualitative, are of considerable importance, given that any degree of residual confounding biasing against lithium suggests that a stronger protective association between lithium treatment and suicide risks exists (to a similar degree) than estimated from logistic regression.

For instance, Kim et al. found that, from VHA charts of patients receiving treatment for depression, suicidal ideation in the past year in the absence of an attempt was associated with suicide with an odds ratio of approximately 3.0.[141] If the sensitivity assumption that diagnostically-coded suicidal ideation underestimates actual suicidal ideation by up to a factor of 6, then the observed imbalance in diagnostically-coded suicidal ideation rates would imply that the overall imbalance in suicidal ideation might account for a bias of up to approximately 0.3 on the observed odds ratio. The impact of this imbalance would be less than this amount if it is assumed that diagnostically-coded suicidal ideation rates underestimate genuine suicidal ideation by a factor less extreme than 6-fold; however, the impact of this imbalance could be greater than approximately 0.3 if it is assumed that some of this suicidal ideation was also associated with suicidal planning or preparatory actions acquiring access to means, both of which are more

strongly associated with suicide risk.[141] Thus, it is plausible that the imbalance in suicidal ideation could account for approximately 60% of the increased risk observed among patients stopping/modifying treatment over 365 days (central estimate OR=1.51), although certainly the impact of suicidal ideation, depending on its prevalence and severity, on observed risk could also be less or more than this amount. Of particular relevance, an impact of this magnitude upon residual confounding resulting from the imbalance of suicidal ideation would imply a central estimate odds ratio during active treatment over 0-365 days of approximately cOR=0.68, rather than the cOR=0.86 that was observed.

The second line of evidence is the fact that central effect estimates did change, albeit modestly, during the modified propensity score sensitivity analysis in which almost half the covariates were removed (Appendix 1-6). This suggests that some degree of confounding amplification may exist, which by extension then implies the presence of some degree of residual confounding still persisting after the propensity score matching. (Some residual confounding must exist for residual confounding amplification to have any noticeable quantitative effect).

The third line of evidence is the least definitive and straightforward, but relates to the observation of increased risks among patients stopping or modifying initial treatment over 0-180 days. Of note, these observed risks both strengthened and remained significant when only patients stopping, rather than modifying or resuming treatment were considered (Table 2-4, Footnote i), when risks were examined among patients stopping or modifying treatment over 91-180 days (Appendix1-4), and among all patients initiating lithium compared to valproate over 91-180 days (Appendix 1-4). Residual confounding is one of several possible explanations for the observation of increased risk in patients discontinuing one treatment compared to discontinuing another treatment[48, 49] However, this conclusion is far from definitive because several other processes can influence risk among patients who have stopped initial treatment. In the strictest sense, for risks in "former users" to most directly reflect baseline confounding, such confounding must not vary substantially over time, substantial differences must not exist in the rates or reasons for discontinuing treatment between treatment groups,[48] any effects

from active treatment must not persist into the period after discontinuation, and/or discontinuation of one medication cannot generate different risks (e.g., "rebound" effects) than discontinuing the comparison medication.

In Appendix 1-5D, we discuss the evidence from the time course of risk in patients discontinuing treatment that suggests to us that at least some of the risk observed in patients who have discontinued treatment is attributable to risks results from discontinuation (or selection), not from confounding. However, in Appendix 1-5E we will discuss an integrative synthesis that includes a consideration of the associations observed in the intent-to-treat sample and among patients still receiving initial treatment from 181-365 days. These associations suggest not only that an association between active lithium treatment and reduced suicide risk is possible, but also that it may be sizable.

Appendix 1-5D. Likelihood of Differential Suicide Risk Associated with Lithium versus Valproate Discontinuation

Several lines of evidence appear to support the possibility that differential risk of suicide may be associated the discontinuation of lithium compared to valproate. If so, such differences could explain (along with residual confounding and possibly some contribution from differences in selection during follow-up) part or all of the statistically significant increased risks associated with patients discontinuing lithium compared to valproate over 0-180 days and, in the survival analysis, over 91-180 days.

The most important line of evidence is that the time course of risk appears to more straightforwardly support the possibility of differential risks upon lithium versus valproate discontinuation than residual confounding. If residual confounding was primarily responsible, the general expectation would be that the time-varying pattern of suicide risk over time among patients discontinuing lithium treatment would decrease progressively from a peak in the first 90 days. A pattern of risk consistent with this possibility has been observed in relation to antidepressant initiation,[17, 122] although the possibility cannot be excluded that antidepressants may increase suicidal behavior risk early in treatment in some sensitive individuals, especially of younger age.[142]

If initiation of a medication is viewed as a clinical event that likely serves as a marker of a patient sufficiently symptomatic to be at higher than usual risk, [18] then it may be relevant to consider the time course of risk concerning other, even more dramatic clinical events which may serve to identify patients as being at particularly high risk (such as suicide attempts and hospital discharges). In these instances, highly time-limited periods (7-30 days) of extreme risk have been observed (i.e., risks of suicide 10-20X greater than what is observed much later
(e.g., 6 months - 1 year subsequently).[63, 65, 66] Thus, the general expectation would be that residual confounding, if present, would be the greatest over 0-90 days and decrease in subsequent periods.

Instead, the difference in risk observed among patients discontinuing lithium compared to valproate is less evident in the first time period (0-90 days) and then becomes much more evident in the subsequent 90 days. This pattern appears more compatible with a developing risk, i.e., a risk that is not initially present but then becomes increasingly present over the first 180 days of treatment. Such an emerging risk fits closely what would be expected from risks among patients discontinuing treatment, in that patients do not start treatment in the status of no longer receiving treatment. Rather, this status must develop over time. Furthermore, the period of highest risk (0-180 days, or more precisely, as the survival analysis suggests, 91-180 days) would incorporate the period of time in which the majority of patients in the cohort would have been discontinued from their treatment for 1-5 months. Interestingly, this corresponds closely to the period previously observed to be of highest risk for mood episode relapse in patients rapidly discontinuing lithium treatment (median time 4.0 +/- 0.7 months), although this information was gathered from patients who had been on lithium maintenance treatment.[80] Although the time course of the development of risk in the

patients discontinuing initial treatment appears very compatible with an emerging risk such as risk associated with discontinuation itself, some caution in interpretation is warranted. The more modest risks observed from 0-90 days in patients discontinuing initial treatment are based only on 5 total suicides, meaning this estimate (which suggests a lower difference in risk among patients discontinuing treatment over the first 90 days than over days 91-180 days) is particularly uncertain.

A second, related line of evidence supports the presence of some degree of differential risk being associated with lithium, compared to valproate, discontinuation. It is important to note that between 0-90 days and 0-180 days the movement in the estimate of risk among patients still receiving initial treatment is modest. A central estimate cOR of 0.87 exists over 0-90 days, compared to a cOR of 1.0 over 0-180 days. (Since there were no suicides in the valproate group on current treatment from 91-180 days, a hazard ratio for this period among patients receiving initial treatment unfortunately is not available. Therefore, comparisons must be made between the nested 0-90 and 0-180 day periods). If confounding was the primary or exclusive explanation for why risks in patients discontinuing treatment increased from 1.49 (central estimate) over 0-90 days to 2.72 over 0-180 days, quite a sizeable increase, then risks associated with active treatment would be expected to rise considerably unless the treatment effect strengthened quickly. Some minor increase in risk associated with active treatment does occur, but nothing similar in size to the increase occurring among patients discontinuing treatment. This suggests that the increase in patients discontinuing lithium treatment most likely not due primarily to increases in confounding. This pattern also suggests the elevated risks are not the product of selection during follow-up favoring the highest risk individuals being discontinued from lithium. While such selection would produce increased risk of suicide being observed in conjunction lithium, rather than valproate, discontinuation, it would also be expected to result in a compensatory *decrease* in risk in patients still receiving active treatment. Instead, the risk increases slightly (from cOR = 0.87 to cOR = 1.0). Thus, the relative stability observed in the estimate of suicide risk associated with the two treatments among patients still receiving initial treatment between 0-90 and 0-180 days suggests that the differing risks of suicide observed upon lithium, compared to valproate, is most easily explained by a process that would be restricted just to the patients discontinuing initial treatment. The two other candidates to influence this effect estimate, confounding and selection during treatment, both would be expected to substantially affect the treatment effect estimate for patients still receiving active treatment as well. Thus, by process of elimination, this data most easily supports the existence of risks being associated directly with the discontinuation of lithium, compared to valproate. However, how much more likely this possibility is than the alternatives of confounding and selection during treatment cannot be determined, and complex combinations of two or three of these processes occurring simultaneously cannot be excluded.

The third line of evidence is that the risk appears to completely resolve by 181-365 days (although random variation could contribute to this finding). If confounding increased over 91-180 days compared to 0-90 days, it seems appear less

plausible this confounding would virtually completely resolve by 181-365 days. However, a rapid resolution of risk is more plausible if the peak risks directly associated with lithium discontinuation were highly time-limited (as has been somewhat observed for the risk of mood episode relapse).[80] As discussed in Appendix 1-4, very few suicides occurred in patients "newly discontinuing" over 181-365 days, thus the risk estimate for patients stopping/modifying initial treatment are particularly influenced by the risk observed among the large majority of patients are counted as having stopped/modified treatment: those who initially discontinued months ago (i.e., over 0-180 days). Previous studies of suicide risks after discontinuation of lithium maintenance treatment have found the increased risk to be clearly time-limited, although the analyses do not address whether the time-limited period of risk is confined to any time period shorter than the first year after discontinuation.[20] However, it is especially notable that the Goodwin et al. 2003 cohort study,[16] which, like this one examined risks starting at the point of treatment initiation, noted that 32% of all the suicides occurring after treatment discontinuation occurred within the first month after discontinuation. Unfortunately, the Goodwin et al study[16] did not report whether this suicide risk differed between lithium, valproate, and/or carbamazepine (perhaps because the low numbers likely would have prevented any statistical significance findings). Nor did their report describe whether these risks occurred in conjunction with discontinuation occurring early or late in treatment.

One caution for this interpretation is that the two previous studies by Yerevanian and colleagues which compared risk of discontinuing lithium and valproate found discontinuation of both medications to be associated with similar and substantial increased risks of suicidal events (attempts or hospitalization for suicidal ideation).[22, 23] One major potential difference between the studies, however, is that our study was focused exclusively on risks observed within one year of initiation, while the Yerevanian studies typically examined lengthy courses of treatment, on average. For example, average follow-up in the 2007 study was approximately 38 months per patient, of which approximately > 90% of this time was accounted for by time receiving medication. This longer follow-up time, including more time on medication further from medication initiation, might have served to strengthen the association of both medications with relatively low rates of suicide during treatment. This in turn may have produced a larger contrast upon discontinuation, especially if the discontinuation is associated with decompensation prompting either the patient or provider to discontinuation treatment.

Appendix 1-5E. Summary and Integration of Key Findings

Although our data initially appears most straightforwardly consistent with an interpretation that lithium treatment either is associated with similar suicide risks or increased suicide risks compared to valproate in this Veterans cohort over the first 365 days of treatment, several important complexities present themselves. The first complexity is that any increased risk associated with lithium treatment appears to be entirely or almost entirely associated with risks observed after treatment discontinuation, not during active treatment. This suggests that, in contrast to most comparative

effectiveness studies, the degree to which the medications may differ in effectiveness relates substantially to what is observed after treatment discontinuation, rather than during treatment.

The second major complexity is the highly time-varying pattern of the intent-to-treat risks, going from a central estimate hazard ratio of 0.95 at 0-90 days to 3.50 at 91-180 days to 0.81 at 181-365 days, with the 91-180 day hazard ratio being statistically significant. This pattern, although it could reflect a large contribution from random variation, appears suggestive of a substantial emergent risk developing after 90 days of treatment in those discontinuing treatment. In general, baseline confounding occurring after a marker of high risk such as treatment initiation[18] would be expected to diminish steadily, with the highest risk being observed shortly after treatment initiation (Appendix 1-5D). Therefore, residual confounding does not appear to be a good candidate to explain this emergent risk, nor does another possibility, selection during follow-up. In theory, selection of patients during follow-up could certainly produce such an emergent risk in patients discontinuing lithium, if the patients being discontinued from lithium were those at particularly high risk for suicide. However, in the absence of a genuine medication effect, selection during follow-up in two treatment groups with similar discontinuation rates would not be expected to alter intent-to-treat risks measured from treatment initiation (Appendix 1-5A). Although we observe only marginally significant intent-to-treat risks from initiation over 180 days, it also should be noted that if selection was occurring in the direction to explain the elevated risks among patients discontinuing lithium at 180 days, this selection should also engender a reduction in risk among patients remaining on initial treatment (since the highest risk individuals are being removed from this patient group). This is not what is observed, instead, the central estimate of the risk among patients still receiving initial treatment stays essentially the same (to be precise, increases slightly, rather than decreases). If confounding or selection during follow-up is not accounting for this sharply increased risk at 91-180 days, then the emergent risk that is suggested is suicide risk that is associated with the discontinuation of lithium early during lithium treatment. Furthermore, it would appear that such risk is somewhat limited to a relatively brief period after lithium discontinuation, seemingly similar to the timing of risks for mood episode recurrences previously noted for others after rapid discontinuation of lithium treatment.[80]

The third major complexity is that intent-to-treat risks do not remain substantially elevated in the final time period, but rather decrease to such an extent that an intent-to-treat estimate is in the direction of lower suicide risk associated with lithium for 181-365 days (albeit this finding is clearly nonsignificant, and thus potentially being the result of chance). Although the role of chance limits the weight that can be placed on this finding, this is potentially a very important observation, given that both any residual confounding and risks associated with discontinuation appear clearly to be most likely associated in the direction of greater suicide risk observed with lithium. If both these important components to an intent-to-treat estimate would be expected to be in the direction of increased risk being associated with lithium treatment, then an obvious candidate that remains to account for a reduction in suicide risks is an active medication effect among

the patients still receiving initial treatment. The observed risks over 181-365 days also suggest that both any confounding and the effects of risk from discontinuation for the bulk of the cohort have resolved, a conclusion consistent with the risk of OR = 0.93 observed among patients who have discontinued treatment.

The opportunity to examine risks from this period (181-365 days after initiation) are of particular interest, not only because of the possibility for observing relatively unconfounded estimates of lithium's treatment effect that the data somewhat suggests, but also because of the potential size of that possible effect. The reduced risk estimated by the central estimate of the intent-to-treat hazard ratio for this time period (0.81) is not clinically insubstantial, although it must be kept in mind that this association does not achieve statistical significant and part of this reduction appears accounted for by the slightly reduced risk of suicide in patients who have discontinued lithium, compared to valproate, treatment (the large majority of which would now would be separated by months from their discontinuation event). Nevertheless, this observation suggests there is at least a reasonable possibility that active lithium treatment is serving to reduce suicide risks of the fraction of patients still receiving lithium within this time period. Given that active treatment now only represents about 16% of the total follow-up time contributing to the intent-to-treat estimate within this time period, to the extent that the effects of active treatment are contributing to this intent-to-treat effect estimate, the association between active lithium treatment and reduced suicide risk could be rather sizeable. Consistent with this inference, a sizable association (cOR = 0.26, 95% CI 0.03, 2.35) is what is observed among patients still receiving their initial treatment, although it does not reach statistical significance (p = 0.23), is informed by just 5 suicides total (1 in lithium recipients and 4 in valproate recipients), and may also reflect contributions from any effects of differential selection during follow-up. Thus, despite the overall nonsignificance of our primary analysis over 0-365 days, and the significant associations observed between lithium discontinuation and increased suicide risk over the first 180 days, our data also suggests, although with much less confidence, that a clinically meaningful reduction in suicide risk may be associated with active lithium treatment after just 181 days of treatment. In this context, it is noteworthy that when the analysis of individuals still receiving initial treatment is continued to 730 days, this association strengthens (cOR = 0.18) and almost achieves marginal statistical significance (p = 0.11) (Appendix 1-4), although this relationship does not persist until the end of follow-up.

It should also be noted that, because of the direction of any residual confounding apparently biases against lithium, it is plausible that similar reductions in suicide risk exist over 0-180 days, but this reduction is simply concealed by confounding. In addition, as mentioned in Appendix 1-5A, it is plausible that if selection early during treatment paralleled initial confounding, then selection during follow-up may have also biased against finding associations between lithium and reduced suicide risk during active treatment. In a sense, this study may provide a useful contrast to earlier literature in which baseline confounding was expected to be in the direction of finding an association between lithium and decreased suicide risk.[14] If selection during follow-up generally paralleled initial confounding in this early literature, associations between

lithium and suicide risk would be expected to overstate lithium's benefits. The extra significance of the direction of baseline confounding in predicting both the direction of bias to the effect estimates from confounding, but also potentially from selection during treatment, reinforces the value of identifying cohorts that appear to have minimal confounding in future nonrandomized studies of lithium and suicide risk.

Given that the likely direction of any residual confounding, and potentially selection during follow-up, biases against observing any association between lithium and reduced suicide risk, this study is more likely to underestimate than overestimate the benefits of active lithium treatment. This aspect of this study should be kept in mind when comparing this study to other literature.

Appendix 1-5F. Recommendations for Clinical Practice Emerging from This Study's Results

This qualitative integration of the evidence from our study supports a number of important clinical and research recommendations. First, the clearest findings from our study related to the statistically significantly increased risk of suicide among patients discontinuing lithium over the first 180 days of the study. Although these risks are in the likely direction of any residual confounding, the distinct time course of their emergence strongly suggest the presence of at least some degree of increased risk being associated with lithium, compared to valproate, discontinuation. These findings indicate that patients should be warned about the possibility of experiencing an increased risk for suicide should they choose to discontinue their treatment, and that providers should also be educated concerning this possible unintended consequence of lithium treatment. In general, persistence with lithium treatment once initiated should be maximized if possible and clinically appropriate. Maximizing persistence may have the dual benefit of maximizing any beneficial associations of active lithium treatment with reduced suicide risk and minimizing the risks associated with lithium discontinuation. Useful reviews of evidence-based approaches to maximize adherence to mood stabilizers have been published.[116] Providers should also be educated that, should discontinuation prove necessary, gradual, rather than rapid, discontinuation of lithium should be implemented when clinically appropriate. Gradual discontinuation appears to substantially reduce the risk of mood episode relapse[79, 80] and thus plausibly may also decrease any associated suicide risk. (However, the possibilities that this difference in risk may relate all or in part to the characteristics of patients able to discontinue gradually versus those not able to discontinue gradually cannot be currently ruled out). Patients who do discontinue treatment should also be educated to monitor themselves closely, and providers should monitor such patients closely when feasible. Such monitoring is already recommended in general after mood stabilizer discontinuation.[78]

In addition, this study provides several important research recommendations, which are discussed further in Appendix 1-7.

Appendix 1-6. Modified Propensity Score Analysis

As part of the evolution of propensity score methods, concerns have been raised that inclusion of variables that are not strongly related to outcome may actually increase the impact of confounders not included in the analysis.[46, 59] Processes for handling this possibility have been debated, but one approach that has been evaluated in the literature has been to restrict the propensity score simply to variables with a strong association with outcome (e.g. +/-20 %).[143] We applied this approach to our data in an exploratory analysis focused on the time period with the most statistically significant findings (0-180 days). All covariates associated with a univariate OR with suicide of between 0.83-1.19 were removed from the propensity score (approximately 50% of the total number of covariates). Because of the interest in risk in patients stopping treatment entirely, rather than simply modifying treatment, for this exploratory analysis we removed patients who modified or resumed their treatment to obtain a "no longer exposed" sample of follow-up time from patients restricted to those who discontinued treatment. The following results were obtained, compared to the results for this analysis for the full propensity score (given in Table 4, Footnote i, and below).

Full Propensity Score Analysis:

Patients Still Receiving Initial Treatment: Conditional Odds Ratio (cOR) = 1.00, 95% CI 0.5, 1.96; Rate Ratio = 1.01

Patients No Longer Exposed (i.e., removing patients who modify or discontinue and subsequently resume treatment): Odds Ratio = 3.61, 95% CI 1.34-9.73, Rate Ratio = 3.60

Modified Propensity Score Analysis (removing variables not associated with a +/-20% change in the odds of suicide from the propensity score):

Patients Still Receiving Initial Treatment: 0-180 days: cOR = 1.00; 95% CI 0.58, 1.72; Rate Ratio = 1.22

Patients No Longer Exposed: 0-180 days: OR = 3.00; 95% CI 1.19, 7.55; Rate Ratio = 2.98

This analysis provides suggestive evidence that the overall contribution of any amplification of confounding to the effect estimates may be relatively modest (given that only a 24% change in risks among patients discontinuing treatment is observed after this substantial change in the propensity score was executed). Such a finding appears consistent with other lines of evidence suggesting that residual confounding amplification does not overly impair this analysis (Appendix 1-5B). The development of methodology for assessing the possibility of confounding amplification, however, is still embryonic.

That consideration, plus the large role that may be played by statistical uncertainty given the wide confidence intervals in this study, means definitive conclusions about the amount of residual confounding amplification cannot be reached. Since propensity score methods are specifically sensitive to this potential effect, the possibility of residual confounding amplification should be kept in mind in the interpretation of this study's results and during comparisons to previous findings.

Appendix 1-7. Suggestions Concerning Future Research

Despite this study's unprecedented size, one fundamental conclusion of the study is that further research concerning the associations between lithium and suicide risk needs to be vigorously conducted. Even without this study, differences between recent and past randomized and nonrandomized research suggest that questions remain concerning the degree to which lithium treatment may be associated with reduction in suicide and suicidal behavior risk. Furthermore, there is an increasing awareness among healthcare researchers that nonrandomized studies may easily contain confounding bias related to the characteristics of those patients, even within specific diagnostic categories, that are chosen to initiate one medication compared to another, despite efforts to rigorously control for measurable patient differences. Mental health research may be particularly sensitive to this potential confounding. As we indicate, despite the extensiveness of our covariates, like virtually all of its predecessors this study lacks extensive information concerning several potential confounders such as suicidal ideation, planning and means, psychiatric symptoms, and recent stressors.

Nevertheless, our study, through its inclusion of intent-to-treat and post-discontinuation risk estimates, will hopefully serve to help focus future mental health research into the question of lithium and suicide risk. First, this study has indicated that increased investigative focus should be placed on examining the possibility that lithium treatment, in cohorts with very high discontinuation rates, might actually increase overall (intent-to-treat) risks of suicide in the short term, if it is determined that lithium discontinuation does indeed pose greater acute suicide risks than valproate discontinuation. Second, however, our results remain compatible with the possibility that active treatment with lithium may be associated with substantial reductions in suicide risk. This also should be the focus of energetic follow-up research, especially since relatively few effective interventions against suicides are known, including among medications. Consider, for instance, that if sufficient residual confounding and/or selection during follow-up biasing against lithium persists in our analysis to conceal a protective association of lithium with suicide even on par with that of clozapine (HR = 0.76) for suicidal events, lithium would likely be a much more valuable intervention. This potential value arises from lithium's potential impacts on suicide mortality (not observed for clozapine[28]), use across a broader range of psychiatric diagnoses, and much less burdensome monitoring requirements.

This study has reemphasized the need to become even more rigorous about attempting to control for confounding at baseline, given that imbalances appear to persist

in the factors (e.g., diagnostically-coded suicidal ideation) not able to be included in the propensity score (despite very tight balance being achieved in numerous other covariates). Such additional research could take several forms. Randomized trials, although challenging to execute, would undoubtedly provide the most rigorous, unconfounded answer regarding the effectiveness of lithium and comparison medications against suicidal behavior, if such trials can be conducted practically (numerous participants would be needed), safely (recommendations how to do so have been advanced),[28, 144] and ethically (i.e., through comparisons with genuine equipoise). Instrumental variables such as prescriber preference variables[129] may also be valuable to investigate, given the potential capability of instrumental variables to balance unmeasured factors. Not all the assumptions underlying instrumental variable analysis, however, can be rigorously tested. Nevertheless, in other treatment studies in which unmeasured confounding was suspected, instrumental variables produce effect estimates closer to those obtained by randomized trials than propensity score methods.[145] Chart review study designs,[146] potentially combined with marginal structural models to address medication, risk factor, and suicidal ideation changes during follow-up, would likely constitute a useful enhancement in cohorts with substantial discontinuation of treatment.

Addressing factors such as suicidal ideation, planning, and means, recent stressors, and recent or current psychiatric symptoms will not be simple, and likely will entail potentially laborious manual chart reviews unless effective automated methods to identify these factors can be developed. It seems likely that case-control or case-cohort designs may need to be adopted to reduce the total number of charts to be reviewed to a feasible number.

In addition, research is needed into how to optimize selection strategies for variables in propensity scores to maximize their benefits in reducing confounding while minimizing potential confounding amplification. This would be especially valuable for studies of suicide risk, since suicide risk is sufficiently multifactorial that approaches such as propensity scores using extensive covariates will likely continue to be desirable. Finally, our study has illustrated the importance of subsequent research adopting methodology such as marginal structural models* that will help facilitate examination of

---

*Because marginal structural models allow a sample to be periodically rebalanced in risk factor composition over time, they may also prove valuable for investigating one further finding from our study. Although not statistically significant and therefore potentially incidental, the finding of consistently lower suicide rates in the valproate cohort after discontinuation of valproate than during valproate treatment (Table 2-3, 0-365 day results for the period after stopping/modifying initial treatment, and Table 4, 0-180 day results for the period after stopping/modifying initial treatment) deserves further investigation given the US FDA labeling warning concerning the possibility of increased suicidal ideation or behavior during anticonvulsant treatment (primarily on the basis of data compiled from patients with epilepsy). Another explanation may be that some of the patients subsequently switched treatment to lithium. If this were the case, it would be another suggestive finding of some association between lithium and reduced suicide risk.

outcomes over longer follow-up periods. Studies over longer follow-up could prove any time-varying components to this risk may have largely resolved by that point. Rebalancing on measured factors at that point (i.e., 6 months or one year after initiation) and starting follow-up might be one approach to particularly limit confounding in useful for two reasons. First, it is possible that the associations between lithium or comparison medications and suicide risk strengthen or weaken over time. Second, if baseline confounding is substantially time-varying, periods later in follow-up may have less confounding, as some of our findings suggest.

However, while chart review approaches might provide improved information about suicidal ideation, psychiatric symptoms, and stressors, such information would almost certainly be still incomplete. For this reason, among nonrandomized studies either instrumental variable analysis (as mentioned above) is likely to be particularly valuable, or cohort studies or nested case control studies from cohorts large enough and with sufficient adherence that substantial numbers of patients continue to receive initial treatment for more than just 6 months - 1 year after initiation. It is plausible that if patients were directed to one medication or the other on the basis of suicide risk initially, nonrandomized studies of lithium and comparison medications. The advantages of a patient sample that might be largely devoid of confounding might outweigh concerns that the results would be the most strictly generalizable for the rather select population of patients who are adherent to the medications for 6 months or more.

Large cohort studies may also provide also valuable information for other reasons besides simply whichever methodological enhancements may be applied. Cohorts with greater adherence in general would provide greater power to detect any reductions in suicide risk associated with active lithium treatment (although power to examine risks in patients discontinuing lithium would then be lessened).[48] Research in some international settings for which lithium treatment remains more routine also may potentially yield lower levels of baseline confounding than studies in the United States, in which only a decided minority of patients receive lithium. As studies become more sophisticated and as sample sizes continue to increase, consideration should be given to incorporating information both about dose and compliance based on additional information besides prescription records (e.g., serum blood levels). Regarding dose, one approach might be requiring a minimum dose for study entry (similar to how we excluded "as needed" use of lithium or valproate, use of lithium or comparison medications below a certain threshold dose might be excluded). More useful might be categorizing doses into a "high dose" (i.e., equal or above the median dose) and "low dose" strata. Such determinations may become complicated as patients shift from one status to another over time, although perhaps this could be reflected in marginal structural models or similar approaches. Ideally, judgments concerning dose should take into account a patients' age (and possibly weight), since lower doses are routinely and appropriately used in older patients. As an extreme, formulas exist to calculate expected lithium serum levels based on renal function and other factors, but it is uncertain how valuable this level of precision may be, given that patient fidelity with dosing recommendations usually cannot be ascertained. Serum blood levels can reflect adherence at certain points, but it is unclear how often that

determination will be made close to a point of clinical interest (i.e., an outcome such as suicide or suicidal behavior). Such a determination may often occur close to at least some of the decisions to modify dose, however. Serum blood level information may have to be used more qualitatively to determine whether patients appear to have histories of good or poor treatment persistence in the study, or perhaps examined most closely in subsamples in which the information is available close to a point of clinical interest.

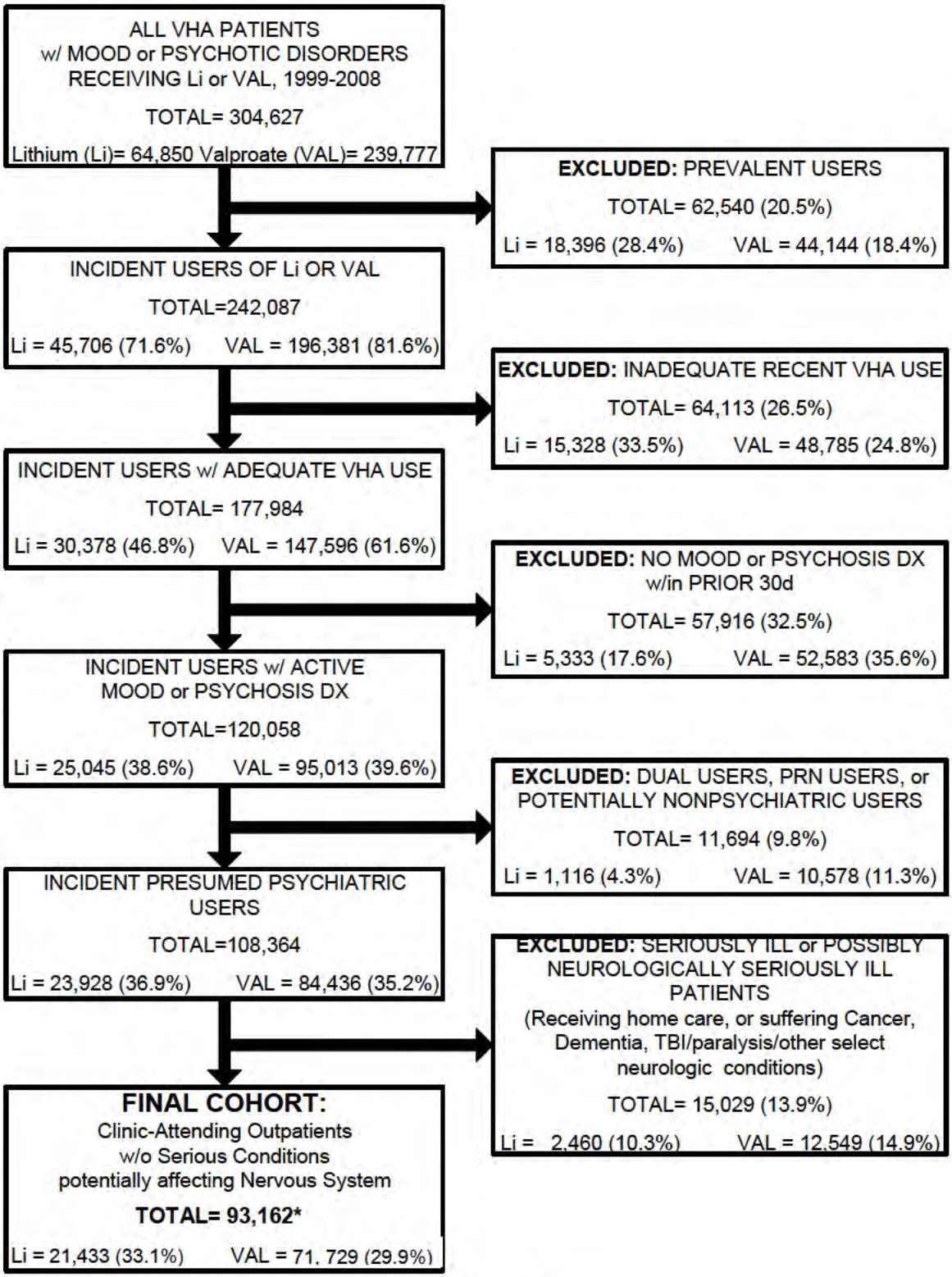
Given the possibility that lithium may increase suicide risks upon discontinuation for some period of time, nonrandomized research should also strive to incorporate an intent-to-treat perspective that ascertains outcomes for individuals both receiving and no longer receiving their initiated treatment. Such an approach will also help facilitate comparisons to randomized research.

It is hoped that the study reported here will help contribute to continued improvements in the investigation of the associations of lithium, and other psychiatric medications, with either decreases or increases in suicide risk. This treatment question is clearly of the utmost importance to patients and providers alike.

Appendix 2

(Additional information of particular relevance to Chapter 3)

Appendix 2-1. Flowchart of Study Cohort Derivation



*excludes 193 patients residing outside the 50 U.S. states

Appendix 2-2. Diagnostic Codes Included in the Cohort

The databases used in this study were initially developed for use in tracking delivered care to a broad collection of VHA patient with depressive or psychotic disorders. Because of this, a considerable range of diagnostic codes were included during database construction. To maximize power and because existing literature suggested than any mortality advantages for lithium might span a variety of diagnoses,[12, 84, 86, 90] we decided to retain this broad group of included diagnoses. Virtually all cohort members had received a diagnosis of bipolar I, II, NOS, Depression NOS, major depressive disorder, schizophrenia, schizoaffective disorder, or "other psychoses" within the past 30 days. Furthermore, the prevalence of these diagnostic categories were highly similar between the two matched treatment groups (i.e., within a standardized difference of < 0.019). Specifically, however, patients could enter the cohort if they had received any one of the following ICD-9 codes in the 30 days prior to lithium or valproate initiation: (more common) 296.0-296.99, 311, and 295.0-295.9, (less common): 297.0-297.3, 297.8-297.9, 298.0-298.4, 298.8, 300.4, 301.12, 309.0-309.1, and 293.83. For more details the interested reader is referred to Appendix 1-1.

Appendix 2-3. Components of Former User Risk

Former user risks have been proposed as a potential indicator of confounding[49] or selection (if confounding has been adequately addressed).[48] However, in actuality differences in former user risks between treatment groups have the potential to be composed of a complex combination of residual confounding, differential selection during treatment, differential discontinuation-associated risks, and any difference in persistent effects from treatment. How often all of these components substantively contribute to former user risk is unclear.

Stated another way, a former user risk (e.g., odds ratio, hazards ratio) of 1.0 is compatible with an absence of confounding, but does not establish this. Similar former user risks cannot establish an absence of confounding if substantial selection during follow-up, discontinuation-associated risks, or persistent effects from active treatment are present. However, former user risks can still have considerable investigative value. As former user risks gets further from a null value (1.0), it is clear that the presence of confounding, selection, and/or persistence and discontinuation effects becomes increasingly likely, while analyses that achieve former user risks of close to 1.0 may or may not have these substantial effects or biases. However, since two or more strong biases or effects can potentially co-occur in opposite directions even if the former user risk approximates 1.0, a lack of strong biases can never be definitively concluded on the basis of former user risks alone.

Appendix 2-4. Aspects of the High-Dimensional Propensity Score Implementation

Our approach generally followed the original high-dimensional propensity score (hdPS) method used by Schneeweiss and colleagues,[41] with the following exceptions:

1) No automated variable construction or selection based on a combined measure of association with exposure and outcome was performed. Instead initial variables were constructed from most entries in a category (e.g., clinic visits, medications, etc.), except those that were the least common (see below). Selection was later imposed for the outcome-focused propensity score based solely on associations with outcome, not with exposure. A few variables were also removed, as described in Appendix 2-6, relating to very specific measures of past mood stabilizer use which appeared likely to potentially act as instrumental variables.
2) Although limited screening for covariate prevalence was done (few variables were included if present in <1% of the sample), no limitation was placed on whether that a covariate needed to be present in 5% of the patient sample as in the original hdPS method.[41] This was because mortality is an infrequent outcome affecting only a small subsample of the cohort, thus even a covariate of low overall prevalence may contribute to a substantial portion of deaths. Some important covariates judged particularly important *a priori* (e.g., current warfarin prescription, cardiac catheterization in the last 180 days, or age ≥ 80 years old) were included even if present in an overall prevalence of 1% or less.
3) For clearly important variables, a more detailed coding of frequency of occurrence was undertaken than just the absence, presence at < median frequency, presence at greater > median frequency, and presence at > 75th percentile frequency used in the original hdPS method.
4) Greater temporal detail than in the original hdPS method was included for some variables by coding several different time periods for hospitalizations, total provider visits, and other general utilization variables, as well coding frequency in two time periods for specific clinic visits (0-180 days and 181-365 days) and medications (current prescriptions and recent, but not current, prescriptions [last days supply ending within the last 180 days]). This strategy was implemented to make information about recent care that might contribute baseline mortality risk less dependent on the exact relationship of the medication initiation date to receipt of medications or services. In addition, this approach might include more detailed information that may be relevant to recent adherence behavior.

Appendix 2-5. Summary of Variables

The following Table summarizes the variables in the initial high-dimensional propensity score:

Appendix 2-5 Supplementary Table 1. Summary of Variables Included in the Initial High-Dimensional Propensity Score (forming the basis of those variables selected for the Outcome-Focused Propensity Score)

Type of Patient Characteristic	Covariates
	General Covariates
Demographics	10 Covariates (49 indicators) including age (11 5-year categories), sex, self-reported Race (6), ethnicity, marital status (4), income(6), Disability Status(4), distance to VA facility (4), urban/rural hospital location, and fiscal year of medication start (11)
Presenting Diagnosis	9 Covariates denoting psychiatric diagnosis in the past 30 days: Bipolar I, Bipolar II, Bipolar NOS, Major Depression, Depression NOS, Schizophrenia, Schizoaffective Disorder, Other Psychoses, and ≥ 2 of these diagnoses
	NonMental Health Covariates
General Utilization	Total number of current prescriptions, possibly discontinued prescriptions (expired in last 30 days) and recently discontinued prescriptions (expired from 31-180 days), Total number of Provider, Specialist, Surgical, ER Visits, and Inpatient Stays (over from 1- 3 different time periods), number of lab visits in last year, receipt of a flu shot in last year
Diagnoses	41 Variables (44 indicators) relating to diagnoses in the past year, including Total Charlson Comorbidity (CCM) conditions (4 levels), 13 individual CCM Categories (MI, CHF, PVD, CEVD, COPD, Conn Tissue Dz, Peptic Ulcer, Mild Liver Dz, Mod/Sev Liver Dz, DM w/o complications, DM w/ complications, Renal Dz, AIDS/HIV), 11 individual Elixhauser comorbidity categories nonredundant with CCM categories (Arrhythmia ,Weight Loss, Coagulopathy, Pulmonary Circulation Dz, HTN, Valve Dz, Neurodegenerative Dz, Hypothyroid, Obesity, Anemia from Blood Loss, Deficiency Anemia), and 16 additional diagnostic categories (e.g. any fracture, hip fracture, neuropathic pain, back pain, internal injuries, open wounds, etc.), and a Recent Smoking indicator combining Tobacco Dependence diagnosis or treatment)

Appendix 2-5 Supplementary Table 1. (continued)

Current Medications	54 covariates representing current prescriptions for specific medications or medication classes, including antiarrhythmic, several antibiotics and antihypertensive classes, warfarin, antiplatelet agents, statins, oral diabetes medications, HepC medications, opiate pain medications, low and high dose aspirin, NSAIDS, acetaminophen, inhalers, GI protectants, and other medications
Recent Medications	55 Covariates representing medication/medication classes prescribed in last 180 days but not active on initiation date. Includes 52 of the 54 the current medication classes, plus vancomycin, antinausea medications, and bandages.
Hospitalizations	41 Covariates denoting both the presence of specific types of VHA discharges in the last two years (e.g., Medical ICU, Surgical ICU, Cardiology, Cardiac Stepdown, Telemetry, General Medicine, 8 types of surgical hospitalizations, etc.). An additional set of variables were constructed to indicate the specific type of discharge which constituted the most recent VHA discharge. Additional covariates addressed how recently the latest VHA hospitalization preceded medication initiation and the presence of any discharges against medical advice in the past year.
Outpatient Providers Visited	156 Covariates (300 indicators) denoting frequency (typically 0/1/2+ visits) of a large variety of outpatient clinics visited in the last 180 days and in days 181-365 prior to initiation. These include specific medical specialties, specific surgical specialties, anticoagulation clinic, outpatient pharmacy consultation, physical therapy, pacemaker and cardiac catheterization clinics, weight loss clinics, and nonmedical specialty services such work therapy and chaplain visits.
Diagnostic Tests	9 covariates (16 indicators) for frequency of tests in prior year, including X-Ray (3), CT/MRI(3), EKG (3), Echocardiogram, Ultrasound, Endoscopy, Nuclear Medicine, PFT, and Angiogram.

Appendix 2-5 Supplementary Table 1. (continued)

Substance Abuse Diagnoses	41 variables diagnoses in the past year of alcohol, amphetamine, cannabis, cocaine, opioid, sedative, stimulant, hallucinogen, and other/unspecified substance categories abuse or dependence. For each substance, 4 variables were constructed reflecting the diagnoses categories of abuse, dependence, remission from abuse, and remission from dependence. In addition, variables were constructed to reflect combined drug dependence (with or without opioids), alcohol intoxication, and alcohol psychosis.
Substance Abuse Treatment	11 variables (19 covariates), including frequency of individual or group substance abuse treatment in last 180 days or days 181-365 prior to initiation, and current or recent prescription of disulfarim, naltrexone, buprenorphine, or methadone.]
Other Psychiatric Covariates	Numerous covariates including General Mental Health Utilization variables, comorbid psychiatric diagnoses, current psychiatric medications, recent psychiatric medications, number, type and timing of recent psychiatric hospitalizations, recent diagnosed suicide attempts, and types of psychiatric outpatient clinic utilization (e.g., psychiatry, psychotherapy, PTSD-focused, etc.)
Aggregate Mortality	5 indicators denoting age and sex-adjusted state mortality risk, derived from CDC data, grouped into 5 categories (approximate quintiles).
VHA Hospital Network Mortality/ Quality of Care	6 indicators denoting categories of rate of risk-adjusted mortality for the VA Integrated Service Network (VISN) where patient received care. (Categorization based on data reported in Reference 16).

Appendix 2-6. Derivation Process for Variables

This appendix is provided to document for interested readers how the additional covariates in the nonsuicide mortality high-dimensional propensity score were derived (the state all-cause mortality rate variables and the VHA VISN risk-adjusted mortality variables). For the remainder of the variables, the reader is directed to Appendix 1-2. These variables form the basis from which the 523 covariates in the outcome-focused propensity score were selected. The selection process for the outcome-focused propensity score variables is also described.

Geographic All-Cause Mortality Risk: Indicator variables were constructed to classify patients into 5 categories (approximate quintiles) of age-adjusted regional (state-level) mortality risk, based on publically available data from the Centers of Disease Control for the years 2000 and 2007.[134] Because these statistics would include the deaths of Veterans occurring in this period, there is the potential for control for "predictors" that include outcome-related information, but this bias is expected to be exceedingly small, given the large number of deaths that occurred across these states over eight years, and our sample accounted for less than 600 nonsuicide deaths over that period. A geographic all-cause mortality indicator was included to guard against the possibility of regional differences in prescribing patterns creating a spurious association between treatment and mortality.

VHA Hospital System (VISN) Mortality Risk: Indicator Variables were used to classify patients into 6 categories of risk-adjusted (age, gender, Charlson Comorbidity Index, perceived physical health, and perceived mental health) all-cause mortality risk, using information from VA surveys administered in 1998 and 1999.[100] Although this information is most accurate for the very beginning of the study period (mid-1999), it was judged that having some indicator of both mortality variation among VA Hospital Systems would be helpful. While the relationship between all-cause mortality and quality of care is controversial, such variables might help limit possible spurious associations between a treatment and mortality due to general tendency for hospitals providing higher- or lower- quality care to have providers who favored one or the other treatment.

Selection of Covariates for the Outcome-Focused Propensity Score: Covariates were selected with an association with mortality of +/- 20 percent (odds ratio of ≥ 1.2 or ≤ 0.83), as has been done previously.[102] Determining whether a dichotomous variable has a 20% association with mortality is generally straightforward, however for covariates with more than two possible levels (e.g., age), determining which variables are included or excluded becomes more of a matter of judgment. Either highly restrictive (requiring all categories of the variable to have an association with mortality of $\geq 20\%$) or highly permissive criteria (requiring only 1 category of the multilevel variable to have an association of $\geq 20\%$) could be envisioned. We adopted a compromise approach in which multilevel variables were included in the outcome-focused propensity score only if

a majority of level of that variable had a +/- 20% odds ratio association with nonsuicide mortality, except for a very few limited exceptions.

Appendix 2-7. Initial Propensity Score Results

Matching upon the initial propensity score produced results that appear to be consistent with some degree of "amplified confounding." [46, 59] For this reason we chose to report the outcome-focused propensity score results throughout the manuscript as likely more unbiased. However, for completeness, we report the initial propensity score results here and compare these results to Tables 3 and 4 of the manuscript.

Comparing the former user values between the initial and outcome-focused propensity score-matched analyses suggests that the outcome-focused propensity score is less confounded than the initial propensity score: At 90 and 365 days the former user hazard ratios are closer to 1.0 for the outcome-focused propensity score than the initial propensity score (central estimate HRs 0-90 day: 0.88 (outcome-focused score) versus 0.67 (initial score); 0-365 days: 1.02 (outcome-focused score) versus 0.84 (initial score). This pattern does not hold for 0-180 days, but in this case the results are consistent with potential discontinuation risks being attenuated by the presence of greater confounding for the initial propensity-score matched cohort in the direction of better outcomes for lithium (central estimate HR = 1.19 versus 1.54 for the outcome-focused score). In addition, the initial propensity score intent-to-treat and former user hazard ratios over 0-90 and 0-365 days are more similar than the outcome-focused propensity score hazard ratios to the hazard ratios observed prior to propensity score matching (Appendix 2-9). This suggests greater residual confounding for the initial propensity score matched cohorts. Since this analysis includes more covariates, this suggests a greater amplification of unmeasured/incompletely measured confounding as some have suggested can occur with control of measured covariates not substantially associated with outcome.[46, 101] Interestingly, the as-treated initial propensity score results are not closer to the unmatched results than the outcome-focused results, a finding that suggests some contribution from random error or that the effects of amplified confounding/less important covariates may warrant more theoretical or empirical investigation.

Also of note, the former user risks of HR = 0.84 over 0-365 days actually exceeds in magnitude the intent-to-treat risk estimate. This same pattern is observed over 0-90 days. This pattern of risk suggests that the initial propensity score intent-to-treat risks may be largely or even initially explained by confounding, and that the former user risks is made up of substantial confounding combined with an additional element (e.g., random error).

APPENDIX 2-7 SUPPLEMENTARY TABLE 1. Risk of Nonsuicide Mortality (Intent-to-Treat Cohort, Initial Propensity Score-Matched)

Time Period	Hazard Ratio (Lithium/Valproate)		
	Intent-to-Treat	During Initial Exposure (As-Treated)	Subsequent Nonexposure (Former User)
0-90 Days	0.72[a] (0.55-0.95)	0.81 (0.57-1.14)	0.67 (0.35-1.25)
0-180 Days	0.97[b] (0.82-1.15)	0.83 (0.61-1.12)	1.19 (0.80-1.77)
0-365 Days	0.87[c] (0.77-0.97)	0.77 (0.57-1.03)	0.84 (0.65-1.07)

Appendix 2-8. Propensity Score Matching Details

Our propensity score matching was performed using greedy-matching involving freely available SAS code from the Mayo Foundation for Medical Education and Research[99] as well as SAS code from Fairies et al., Chapter 3 of the SAS Press book "Analysis of Observational Health Care Data Using SAS."[61]

Because it is not part of this published code, we did not trim our propensity score cohorts to a "Common Support Area" prior to matching. Perhaps due to the large preponderance of patients initiating valproate compared to lithium and the wide, overlapping propensity score distribution for both medications, very few patients fell outside the "Common Support Area." This is reflected by the fact that use of fairly standard 0.2 propensity score logit calipers resulted in a narrow propensity score range while including virtually all lithium-treated patients. We more precisely established this for our highly similar analysis of suicide mortality. This analysis also involved nearly complete matching of lithium-treated patients, and these patients were matched using a propensity score that included 98% of the covariates included in the initial propensity score for this study. For this similar analysis, we established that exceedingly few patients fell outside of a Common Support Area (only 0.05% [lithium-treated patients] to 0.12% [valproate-treated patients] of the entire unmatched cohort).

Appendix 2-9. Mortality by Treatment in the Unmatched Cohort and Implications for Confounding

Appendix 2-9 Supplementary Table 1. Risk of Nonsuicide Mortality over 365 days, by Treatment

	Hazard Ratio (Lithium Versus Valproate) (95% Confidence Interval)		
		Stratified by Exposure Status	
Analysis (incident Users)	Intent-to-Treat Sample	During Initial Exposure (As-Treated)	During Subsequent Nonexposure (Former Users)
Unmatched Cohort	0.74[a] (0.65-0.84)	0.58[a] (0.45-0.74)	0.77[a] (0.63-0.94)
Outcome-Focused High-Dimensional Propensity-Score Matched Cohort	0.92[a] (0.82-1.04)	0.59[a] (0.53-0.97)	1.02[a] (0.79-1.32)

[a] P values are: p<0.0001 (Intent-to-Treat), p<0.0001 (Current Users); p=0.011 (Former Users)
[b] P values are: p=0.173 (Intent-to-Treat); p=.028 (Current Users); p=0.888 (Former Users)

Of note, for each effect estimate the outcome-focused high-dimensional propensity-score matched cohort produced estimates in the direction of reducing the effect sizes from the unmatched analysis each of which indicated a stronger association with worsened outcomes among patients initiating valproate. (However, this change in estimates between the unmatched and matched analyses was minimal for the as-treated effect estimate). These findings suggest confounding in the overall cohort is in the direction of patients who are less medically ill preferentially receiving lithium (i.e., patients having less risk of nonsuicide mortality at baseline, prior to treatment initiation). As a result, the unmatched associations show stronger effect sizes favoring lithium treatment than the matched analysis.

Appendix 2-10. Persistence with Treatment at 90, 180, and 365 days

Appendix 2-10 Supplementary Table 1. Censoring of Patient Cohorts at 90, 180 and 365 days (Outcome-focused Propensity Score Matched Cohort)

Treatment Status	0-90 Day Follow-up				0-180 Day Follow-up				0-365 Day Follow-up			
	Lithium		Valproate		Lithium		Valproate		Lithium		Valproate	
	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
Still Receiving Initial Treatment	9967	46.8	9852	46.3	5012	23.5	5086	23.9	1621	7.6	1723	8.1
Discontinued Initial Treatment	10501	48.7	11046	51.3	15220	71.5	15699	73.8	18445	86.7	18974	89.1
Initiated opposite mood stabilizer[b]	773	3.6	321	1.5	987	4.6	415	2.0	1133	5.3	470	2.2
Suicide Mortality	15	0.07	15	0.07	17	0.08	16	0.08	18	0.08	20	0.09
Nonsuicide Mortality[c]	32	0.15	54	0.25	52	0.24	72	0.34	71	0.33	101	0.47

[a] n= 21288 propensity-score matched pairs

[b] i.e., switched directly from lithium to valproate or directly from valproate to lithium

[c] Because this Table row reflect patients censored due to nonsuicide death occurring before any treatment impersistence, they are equivalent to the "as-treated" counts.

Appendix 2-11. Numerical Illustration Concerning Possible Residual Confounding

As an example, at 180 days the intent-to-treat HR (central estimate) of 0.97 suggests that residual confounding of only approximately 3-4% lower baseline hazard of mortality at baseline (prior to medication initiation) among patients initiating lithium would be sufficient to yield a central estimate of increased, rather than decreased, mortality risk among all lithium initiators over the first 180 days of treatment. Slightly more confounding would be necessary to yield a central estimate of increased risk over 0-365 days. Statically, the confidence intervals already preclude definitive conclusions concerning net harms or benefits of lithium compared to valproate for these periods. Notably, the intent-to-treat associations over 0-180 days includes the period of strong, statistical significant intent-to-treat associations between lithium initiation and lower mortality risks from 0-90 days.

Appendix 2-12. Expected Decrease over Time of Intent-to-Treat Effect Sizes

Perhaps nonintuitively, if a genuine medication effect exists during active treatment, it can be generally expected that intent-to-treat estimates will predictably weaken over time in a well-controlled analysis. This is because an increasing number of "as-treated" individuals can be expected to discontinue medication treatment over time. As a result individuals "no longer exposed" to the study intervention constitute a greater and greater percentage of total follow-up time.

However, an important point is worth noting. First, as more and more cohort members become nonexposed, their risk actually converges on an intent-to-treat estimate that reflects confounding (as well as any discontinuation-associated risks and random error), not necessarily a true null value (HR=1.0). For this reason, intent-to-treat estimates may not always weaken over time as treatment persistence decreases (for example, in this study, the 0-365-day HR = 0.92, 95% CI 0.82, 1.04]) is further from 1.0 than the 0-180-day HR = 0.97, 95% CI 0.82, 1.15). Thus, the possibility of residual confounding should be considered whenever a pattern of weakening intent-to-treat estimates over time is not observed. While the differences observed in this study are within the realm of random error, it should be remembered that the two hazard ratios are not estimating the same quantity; because, proportionally, considerably less as-treated and considerably more former user follow-up time is present over 0-365 days than 0-180 days. Thus, the expectation would be that the 0-365 day intent-to-treat central estimate effect size would show even a smaller association favoring lithium, not a slightly large one. The fact that this effect estimate reverses in direction from the 0-180 day central estimate, even though substantial random error is present in the comparison, may be notable.

From a related perspective, the rapid increase in central estimate of the hazard ratio from 0-90 days to 0-180 days implies an actual increased risk among patients who previously initiated lithium during the period of 91-180 days. Interestingly, this is the

general pattern of risk we observed in a related study of suicide mortality drawn from the same unmatched cohort (Chapter 2). Our secondary analysis of former user risk suggests a possible reason for this increased risk associated with lithium initiation: a sufficient number of former users experiencing some substantial risk associated with lithium discontinuation to potentially outweigh any benefits experienced by the minority of patients who remain on initial lithium treatment during this period.

Appendix 2-13. Evidence for and against Selection During Treatment

On theoretical grounds, selection during treatment could plausibly explain the reduced risks in as treated patients and enhanced risk among patients discontinuing lithium. This would occur if a greater number of medically ill patients had their lithium treatment stopped because of their deteriorating condition than medically ill patients receiving valproate. In essence, this differential selection would serve to transfer a greater number of high-risk patients receiving lithium treatment than receiving valproate treatment to the category of individuals having discontinued treatment. That is, a greater number of lithium-treated patients change status from being counted as "as-treated" individuals to being counted as "former users."

However, this phenomenon, unless the selection was based primarily on adverse medical risks caused by the medications themselves, would not easily explain the changes in intent-to-treat risks, nor the consistency of the as-treated risks observed. Finally, this possibility is also rendered less likely by the very similar rates of medication discontinuation observed between the treatment groups receiving the two medications over time. However, as others have pointed out, this line of reasoning is not firmly conclusive since patients may discontinue medications at the same rate but for different reasons.[48]

Furthermore, a simple model combining confounding and selection (i.e., positing no medication effects on either risks during active treatment or after discontinuation) does not appear sufficient, since in such a model, in order to explain the intent-to-treat findings, confounding would have to change direction from 0-90 and 91-180 days, and then change direction again, to explain the 0-90, 0-180, and 0-365 day intent-to-treat estimates observed.

It is possible, of course, that random variation does contribute to the 90, 180 and 365 day estimates and perhaps enhances the differences between them, producing spurious changes in direction of the estimates. However, the probability random error explains the difference between the 0-90, 0-180, and the 0-365 day intent-to-treat estimates entirely (i.e., in the absence of residual confounding or genuine medication effects during treatment or upon discontinuation), or to the differences between the 0-90, 0-180 and 0-365 day former user risks entirely, would be considerably less than 50%.

Thus, the simplest consistent interpretation of the outcome-focused results is that some level of overall residual confounding biasing against valproate persists in the 0-365 day analyses, although it is not the only contributor to the risk estimates. When

dissociation-associated risks and/or (less likely) selection effects weaken from 0-180 days to 0-365 days, any confounding might then serve to "pull" the former user risk much closer to the null over a relatively short period, and make the intent-to-treat estimates more negative.

Appendix 2-14. The Value of Lithium Treatment Persistence

A corollary exists to the interpretation given above that some level of residual confounding is likely in the analysis is biasing towards lower mortality risk among patients initiating lithium. That corollary is that the assessment of whether patients initiating lithium are at greater or lesser mortality risk when all associations (active treatment and dissociation-associated) are considered is made more complicated. This is especially true because the margin of beneficial association observed at 0-180 and 0-365 days is generally small. That is, as pointed out in Appendix 2-11, even small to relatively small levels of residual confounding (for the 0-180 day and 0-365 day analyses, respectively) would be sufficient to conceal any overall hazardous treatment effects associated with lithium compared to valproate. For this reason, we note in the manuscript that further research is clearly needed and caution should be exercised regarding any judgments of whether greater or lesser lithium use would be desirable.

However, the recommendation for increasing persistence with lithium treatment is clearly indicated by the data, regardless of this uncertainty. Whether lithium is reducing mortality risks (relative to valproate) during active treatment, or increasing mortality risks upon discontinuation, or both, in any of these scenarios increasing persistence with lithium treatment should result in mortality benefits. Put another way, regardless of whether the net impact of lithium over the entire follow-up period leads to lesser or greater mortality risks than initiation of valproate, once the decisions has been made to initiate lithium, boosting persistence with lithium would be indicated and beneficial.* This is because the study suggests that lithium has at worst a neutral effect during active treatment and potentially a beneficial effect, whereas any hazards from lithium treatment relative to valproate appear to occur upon discontinuation. Emphasizing treatment persistence would have the effect of increasing any benefits experienced during active treatment and reducing any risks resulting from discontinuation. Regardless of which

---

* One possible exception to this principle exists. If confounding biasing against valproate is so great that a net hazard is associated with even active lithium treatment, then depending on this magnitude of this hazard, it may be worth even in the short term the patient discontinuing lithium and suffering potential short-term discontinuation hazards. While our data, being observational (nonrandomized) cannot exclude this possibility, it is not substantially supported from the study findings.

risk predominates, once a decision is made to initiate lithium, efforts to boost treatment persistence when feasible appear likely to benefit the patient.

Appendix 2-15. Challenges in Completely Modeling Important Risk Factors

It is important to recognize there is inherent difficult in capturing a fully desirable amount of information concerning some types of variables. Hospitalizations prior to medication initiation are an example. We chose to model these hospitalizations in three ways: multiple indicators for the overall number of recent nonmental health hospitalizations, whether any hospitalization of a particular type (e.g., ICU, cardiac, etc.) had occurred in the past 2 years, and what type of hospitalization had occurred most recently prior to medication initiation. However, it can be easily conceptualized that near-complete modeling of hospitalizations experienced by the patient in the last two years might have included the timing and number of days preceding medication initiation for every hospitalization type, and potentially length of stay as well. Furthermore, indicators concerning whether multiple hospitalizations of the same type had occurred and how separated in time the repeat admission were might be desirable. And of course, as pointed out in the manuscript, hospitalizations may occur outside the VHA system for which we have no information. For these types of variables, at some point practical decisions must be made concerning what detail in modeling is appropriate, along with the realization that incorporation into the model of complete information likely can never be achieved.

References

1. Blow FC, Bohnert AS, Ilgen MA, Ignacio R, McCarthy JF, Valenstein MM, et al. Suicide mortality among patients treated by the Veterans Health Administration from 2000 to 2007. Am J Public Health. 2012; **102 Suppl 1**: S98-104.
2. Fontaine KR, Heo M, Harrigan EP, Shear CL, Lakshminarayanan M, Casey DE, et al. Estimating the consequences of anti-psychotic induced weight gain on health and mortality rate. Psychiatry Res. 2001; **101**(3): 277-88.
3. Mann JJ, Apter A, Bertolote J, Beautrais A, Currier D, Haas A, et al. Suicide prevention strategies: a systematic review. JAMA. 2005; **294**(16): 2064-74.
4. Daigle MS. Suicide prevention through means restriction: assessing the risk of substitution. A critical review and synthesis. Accid Anal Prev. 2005; **37**(4): 625-32.
5. Sinyor M, Levitt AJ. Effect of a barrier at Bloor Street Viaduct on suicide rates in Toronto: natural experiment. BMJ. 2010; **341**: c2884.
6. Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. Arch Gen Psychiatry. 2006; **63**(7): 757-66.
7. Brown GK, Ten Have T, Henriques GR, Xie SX, Hollander JE, Beck AT. Cognitive therapy for the prevention of suicide attempts: a randomized controlled trial. JAMA. 2005; **294**(5): 563-70.
8. Hesdorffer DC, Kanner AM. The FDA alert on suicidality and antiepileptic drugs: Fire or false alarm? Epilepsia. 2009; **50**(5): 978-86.
9. Rosack J. FDA Orders Stricter Suicide Warnings for Antidepressants. Psychiatr News. 2004; **39**(8): 1.
10. Tondo L, Jamison KR, Baldessarini RJ. Effect of lithium maintenance on suicidal behavior in major mood disorders. Ann N Y Acad Sci. 1997; **836**: 339-51.
11. American Psychiatric Association. Practice Guideline for the Treatment of Patients with Bipolar Disorder, 2 ed.; 2002.
12. Cipriani A, Pretty H, Hawton K, Geddes JR. Lithium in the prevention of suicidal behavior and all-cause mortality in patients with mood disorders: a systematic review of randomized trials. Am J Psychiatry. 2005; **162**(10): 1805-19.
13. Nilsson A. Lithium therapy and suicide risk. J Clin Psychiatry. 1999; **60 Suppl 2**: 85-8; discussion 111-6.
14. Ernst CL, Goldberg JF. Antisuicide properties of psychotropic drugs: a critical review. Harv Rev Psychiatry. 2004; **12**(1): 14-41.
15. Baldessarini RJ, Tondo L. Lithium and suicidal risk. Bipolar Disord. 2008; **10**(1): 114-5.
16. Goodwin FK, Fireman B, Simon GE, Hunkeler EM, Lee J, Revicki D. Suicide risk in bipolar disorder during treatment with lithium and divalproex. JAMA. 2003; **290**(11): 1467-73.

17. Jick H, Kaye JA, Jick SS. Antidepressants and the risk of suicidal behaviors. JAMA. 2004; **292**(3): 338-43.
18. Valenstein M, Kim HM, Ganoczy D, McCarthy JF, Zivin K, Austin KL, et al. Higher-risk periods for suicide among VA patients receiving depression treatment: prioritizing suicide prevention efforts. J Affect Disord. 2009; **112**(1-3): 50-8.
19. Tondo L, Baldessarini RJ, Hennen J, Floris G, Silvetti F, Tohen M. Lithium treatment and risk of suicidal behavior in bipolar disorder patients. J Clin Psychiatry. 1998; **59**(8): 405-14.
20. Baldessarini RJ, Tondo L, Hennen J. Effects of lithium treatment and its discontinuation on suicidal behavior in bipolar manic-depressive disorders. J Clin Psychiatry. 1999; **60 Suppl 2**: 77-84; discussion 111-6.
21. Muller-Oerlinghausen B, Muser-Causemann B, Volk J. Suicides and parasuicides in a high-risk patient group on and off lithium long-term medication. J Affect Disord. 1992; **25**(4): 261-9.
22. Yerevanian BI, Koek RJ, Mintz J. Lithium, anticonvulsants and suicidal behavior in bipolar disorder. J Affect Disord. 2003; **73**(3): 223-8.
23. Yerevanian BI, Koek RJ, Mintz J. Bipolar pharmacotherapy and suicidal behavior. Part I: Lithium, divalproex and carbamazepine. J Affect Disord. 2007; **103**(1-3): 5-11.
24. Cipriani A, Pretty H, Hawton K, Geddes JR. Lithium in the prevention of suicidal behavior and all-cause mortality in patients with mood disorders: a systematic review of randomized trials. Am J Psychiatry. 2005; **162**(10): 1805-19.
25. Cipriani A, Hawton K, Stockton S, Geddes JR. Lithium in the prevention of suicide in mood disorders: updated systematic review and meta-analysis. BMJ. 2013; **346**: f3646.
26. Oquendo MA, Galfalvy HC, Currier D, Grunebaum MF, Sher L, Sullivan GM, et al. Treatment of suicide attempters with bipolar disorder: a randomized clinical trial comparing lithium and valproate in the prevention of suicidal behavior. The American journal of psychiatry. 2011; **168**(10): 1050-6.
27. Geddes JR, Goodwin GM, Rendell J, Azorin JM, Cipriani A, Ostacher MJ, et al. Lithium plus valproate combination therapy versus monotherapy for relapse prevention in bipolar I disorder (BALANCE): a randomised open-label trial. Lancet. 2010; **375**(9712): 385-95.
28. Meltzer HY, Alphs L, Green AI, Altamura AC, Anand R, Bertoldi A, et al. Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). Arch Gen Psychiatry. 2003; **60**(1): 82-91.
29. Baldessarini RJ, Tondo L, Davis P, Pompili M, Goodwin FK, Hennen J. Decreased risk of suicides and attempts during long-term lithium treatment: a meta-analytic review. Bipolar Disord. 2006; **8**(5 Pt 2): 625-39.
30. Nocente-McGrath C, McIsaac R, Ernst SG. Altered cell fate in LiCl-treated sea urchin embryos. Dev Biol. 1991; **147**(2): 445-50.
31. Manji HK, Bebchuk JM, Moore GJ, Glitz D, Hasanat KA, Chen G. Modulation of CNS signal transduction pathways and gene expression by mood-stabilizing agents:

therapeutic implications. J Clin Psychiatry. 1999; **60 Suppl 2**: 27-39; discussion 40-1, 113-6.
32. Carmen J, Okafor K, Ike E. The effects of lithium therapy on leukocytes: a 1-year follow-up study. J Natl Med Assoc. 1993; **85**(4): 301-3.
33. Colton CW, Manderscheid RW. Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. Preventing chronic disease. 2006; **3**(2): A42.
34. Manji HK, Moore GJ, Chen G. Clinical and preclinical evidence for the neurotrophic effects of mood stabilizers: implications for the pathophysiology and treatment of manic-depressive illness. Biol Psychiatry. 2000; **48**(8): 740-54.
35. Hicks D. Lithium induced renal toxicity--a review of the literature. S D J Med. 1991; **44**(12): 343-5.
36. Gitlin M. Lithium and the kidney: an updated review. Drug Saf. 1999; **20**(3): 231-43.
37. Bocchetta A, Loviselli A. Lithium treatment and thyroid abnormalities. Clin Pract Epidemol Ment Health. 2006; **2**: 23.
38. Weintraub M, Hes JP, Rotmensch HH, Soferman G, Liron M. Extreme sinus bradycardia associated with lithium therapy. Isr J Med Sci. 1983; **19**(4): 353-5.
39. Montalescot G, Levy Y, Hatt PY. Serious sinus node dysfunction caused by therapeutic doses of lithium. Int J Cardiol. 1984; **5**(1): 94-6.
40. Wolf ME, Ranade V, Molnar J, Somberg J, Mosnaim AD. Hypercalcemia, arrhythmia, and mood stabilizers. J Clin Psychopharmacol. 2000; **20**(2): 260-4.
41. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology. 2009; **20**(4): 512-22.
42. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med. 2007; **26**(4): 734-53.
43. Bhattacharya J, Vogt W. Do instrumental variables belong in propensity scores?: Cambridge, MA: National Bureau of Economic Research, 2007.
44. Wooldridge J. Should instrumental variables be used as matching variables? . East Lansing, MI: Michigan State University; 2009.
45. Pearl J. On a class of bias-amplifying variables that endanger effect estimates. Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010); Corvallis, OR: Association for Uncertainty in Artificial Intelligence. p. 2010: 425-432.
46. Pearl J. Invited commentary: understanding bias amplification. Am J Epidemiol. 2011; **174**(11): 1223-7; discussion pg 8-9.
47. Brooks JM, Ohsfeldt RL. Squeezing the balloon: propensity scores and unmeasured covariate balance. Health Serv Res. 2013; **48**(4): 1487-507.
48. Hernan MA, Robins JM. Authors' response, part I: observational studies analyzed like randomized experiments: best of both worlds. Epidemiology. 2008; **19**(6): 789-92.

49. Cooper WO, Habel LA, Sox CM, Chan KA, Arbogast PG, Cheetham TC, et al. ADHD drugs and serious cardiovascular events in children and young adults. N Engl J Med. 2011; **365**(20): 1896-904.
50. Ray WA. Evaluating medication effects outside of clinical trials: New-user designs. Am J Epidemiol. 2003; **158**(9): 915-20.
51. Lauterbach E, Felber W, Muller-Oerlinghausen B, Ahrens B, Bronisch T, Meyer T, et al. Adjunctive lithium treatment in the prevention of suicidal behaviour in depressive disorders: a randomised, placebo-controlled, 1-year trial. Acta Psychiatr Scand. 2008; **118**(6): 469-79.
52. Coryell W, Arndt S, Turvey C, Endicott J, Solomon D, Mueller T, et al. Lithium and suicidal behavior in major affective disorder: a case-control study. Acta Psychiatr Scand. 2001; **104**(3): 193-7.
53. Oquendo MA, Galfalvy HC, Currier D, Grunebaum MF, Sher L, Sullivan GM, et al. Treatment of suicide attempters with bipolar disorder: a randomized clinical trial comparing lithium and valproate in the prevention of suicidal behavior. Am J Psychiatry. 2011; **168**(10): 1050-6.
54. Blow FC, Valenstein M, Austin K, Khanuja K, McCarthy JF. Specialty Care for Veterans with Depression in the VHA: 2002 National Depression Registry Report. Ann Arbor, MI: VA National Serious Mental Illness Treatment Research & Evaluation Center (SMITREC), VHA Health Services Research & Development; 2003 2003.
55. Guzzetta F, Tondo L, Centorrino F, Baldessarini RJ. Lithium treatment reduces suicide risk in recurrent major depressive disorder. J Clin Psychiatry. 2007; **68**(3): 380-3.
56. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major US mortality databases. Ann Epidemiol. 2002; **12**(7): 462-8.
57. McCarthy JF, Valenstein M, Kim HM, Ilgen M, Zivin K, Blow FC. Suicide mortality among patients receiving care in the veterans health administration health system. Am J Epidemiol. 2009; **169**(8): 1033-8.
58. Patorno E, Bohn RL, Wahl PM, Avorn J, Patrick AR, Liu J, et al. Anticonvulsant medications and the risk of suicide, attempted suicide, or violent death. JAMA. 2010; **303**(14): 1401-9.
59. Brooks JM, Ohsfeldt RL. Squeezing the Balloon: Propensity Scores and Unmeasured Covariate Balance. Health Serv Res. 2012.
60. Austin PC. Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. Biom J. 2009; **51**(1): 171-84.
61. Faries DE, Leon, A C, Haro, J M, Obenchain, R L, & SAS Institute. Analysis of observational health care data using SAS. Cary, NC: SAS Institute; 2010.
62. Angst J, Angst F, Gerber-Werder R, Gamma A. Suicide in 406 mood-disorder patients with and without long-term medication: a 40 to 44 years' follow-up. Arch Suicide Res. 2005; **9**(3): 279-300.
63. Goldacre M, Seagroatt V, Hawton K. Suicide after discharge from psychiatric inpatient care. Lancet. 1993; **342**(8866): 283-6.

64. Harris EC, Barraclough B. Suicide as an outcome for mental disorders. A meta-analysis. The British journal of psychiatry : the journal of mental science. 1997; **170**: 205-28.
65. Haukka J, Suominen K, Partonen T, Lonnqvist J. Determinants and outcomes of serious attempted suicide: a nationwide study in Finland, 1996-2003. Am J Epidemiol. 2008; **167**(10): 1155-63.
66. Qin P, Agerbo E, Mortensen PB. Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: a national register-based study of all suicides in Denmark, 1981-1997. Am J Psychiatry. 2003; **160**(4): 765-72.
67. Ahrens B, Muller-Oerlinghausen B, Grof P. Length of lithium treatment needed to eliminate the high mortality of affective disorders. Br J Psychiatry Suppl. 1993; (21): 27-9.
68. Ahrens B, Grof P, Moller HJ, Muller-Oerlinghausen B, Wolf T. Extended survival of patients on long-term lithium treatment. Can J Psychiatry. 1995; **40**(5): 241-6.
69. Muller-Oerlinghausen B, Wolf T, Ahrens B, Schou M, Grof E, Grof P, et al. Mortality during initial and during later lithium treatment. A collaborative study by the International Group for the Study of Lithium-treated Patients. Acta Psychiatr Scand. 1994; **90**(4): 295-7.
70. Yerevanian BI, Koek RJ, Mintz J, Akiskal HS. Bipolar pharmacotherapy and suicidal behavior Part 2. The impact of antidepressants. J Affect Disord. 2007; **103**(1-3): 13-21.
71. Yerevanian BI, Koek RJ, Mintz J. Bipolar pharmacotherapy and suicidal behavior Part 3: impact of antipsychotics. J Affect Disord. 2007; **103**(1-3): 23-8.
72. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. Am J Epidemiol. 2011; **174**(11): 1213-22.
73. Koek RJ, Yerevanian BI, Mintz J. Subtypes of antipsychotics and suicidal behavior in bipolar disorder. J Affect Disord. 2012; **143**(1-3): 27-33.
74. Sondergard L, Lopez AG, Andersen PK, Kessing LV. Mood-stabilizing pharmacological treatment in bipolar disorders and risk of suicide. Bipolar Disord. 2008; **10**(1): 87-94.
75. Smith EG, Sondergard L, Lopez AG, Andersen PK, Kessing LV. Association between consistent purchase of anticonvulsants or lithium and suicide risk: A longitudinal cohort study from Denmark, 1995-2001. J Affect Disord[epub in advance of print]. 2009.
76. Collins JC, McFarland BH. Divalproex, lithium and suicide among Medicaid patients with bipolar disorder. J Affect Disord. 2008; **107**(1-3): 23-8.
77. Gibbons RD, Hur K, Brown CH, Mann JJ. Relationship between antiepileptic drugs and suicide attempts in patients with bipolar disorder. Arch Gen Psychiatry. 2009; **66**(12): 1354-60.
78. Goodwin GM. Evidence-based guidelines for treating bipolar disorder: revised second edition--recommendations from the British Association for Psychopharmacology. J Psychopharmacol (Oxf). 2009; **23**(4): 346-88.

79. Faedda GL, Tondo L, Baldessarini RJ, Suppes T, Tohen M. Outcome after rapid vs gradual discontinuation of lithium treatment in bipolar disorders. Arch Gen Psychiatry. 1993; **50**(6): 448-55.
80. Baldessarini RJ, Tondo L, Faedda GL, Suppes TR, Floris G, Rudas N. Effects of the rate of discontinuing lithium maintenance treatment in bipolar disorders. J Clin Psychiatry. 1996; **57**(10): 441-8.
81. Morton WA, Sonne SC, Lydiard RB. Lithium side effects in the medically ill. Int J Psychiatry Med. 1993; **23**(4): 357-82.
82. Frost RE, Messiha FS. Clinical uses of lithium salts. Brain Res Bull. 1983; **11**(2): 219-31.
83. Norton B, Whalley LJ. Mortality of a lithium-treated population. Br J Psychiatry. 1984; **145**: 277-82.
84. Coppen A, Standish-Barry H, Bailey J, Houston G, Silcocks P, Hermon C. Does lithium reduce the mortality of recurrent mood disorders? J Affect Disord. 1991; **23**(1): 1-7.
85. Vestergaard P, Aagaard J. Five-year mortality in lithium-treated manic-depressive patients. J Affect Disord. 1991; **21**(1): 33-8.
86. Muller-Oerlinghausen B, Ahrens B, Grof E, Grof P, Lenz G, Schou M, et al. The effect of long-term lithium treatment on the mortality of patients with manic-depressive and schizoaffective illness. Acta Psychiatr Scand. 1992; **86**(3): 218-22.
87. Ahrens B, Muller-Oerlinghausen B, Schou M, Wolf T, Alda M, Grof E, et al. Excess cardiovascular and suicide mortality of affective disorders may be reduced by lithium prophylaxis. J Affect Disord. 1995; **33**(2): 67-75.
88. Nilsson A. Mortality in recurrent mood disorders during periods on and off lithium. A complete population study in 362 patients. Pharmacopsychiatry. 1995; **28**(1): 8-13.
89. Brodersen A, Licht RW, Vestergaard P, Olesen AV, Mortensen PB. Sixteen-year mortality in patients with affective disorder commenced on lithium. Br J Psychiatry. 2000; **176**: 429-33.
90. Angst F, Stassen HH, Clayton PJ, Angst J. Mortality of patients with mood disorders: follow-up over 34-38 years. J Affect Disord. 2002; **68**(2-3): 167-81.
91. Kilbourne AM, Morden NE, Austin K, Ilgen M, McCarthy JF, Dalack G, et al. Excess heart-disease-related mortality in a national study of patients with mental disorders: identifying modifiable risk factors. Gen Hosp Psychiatry. 2009; **31**(6): 555-63.
92. Harris EC, Barraclough B. Excess mortality of mental disorder. The British journal of psychiatry : the journal of mental science. 1998; **173**: 11-53.
93. Miller BJ, Paschall CB, 3rd, Svendsen DP. Mortality and medical comorbidity among patients with serious mental illness. Psychiatr Serv. 2006; **57**(10): 1482-7.
94. Zivin K, Ilgen MA, Pfeiffer PN, Welsh DE, McCarthy J, Valenstein M, et al. Early mortality and years of potential life lost among veterans affairs patients with depression. Psychiatr Serv. 2012; **63**(8): 823-6.
95. Young AH, Hammond JM. Lithium in mood disorders: increasing evidence base, declining use? Br J Psychiatry. 2007; **191**: 474-6.

96. Blanco C, Laje G, Olfson M, Marcus SC, Pincus HA. Trends in the treatment of bipolar disorder by outpatient psychiatrists. Am J Psychiatry. 2002; **159**(6): 1005-10.
97. Wolfsperger M, Greil W, Rossler W, Grohmann R. Pharmacological treatment of acute mania in psychiatric in-patients between 1994 and 2004. J Affect Disord. 2007; **99**(1-3): 9-17.
98. Shulman KI, Rochon P, Sykora K, Anderson G, Mamdani M, Bronskill S, et al. Changing prescription patterns for lithium and valproic acid in old age: shifting practice without evidence. BMJ. 2003; **326**(7396): 960-1.
99. Wyss R, Girman CJ, Locasale RJ, Alan Brookhart M, Sturmer T. Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study. Pharmacoepidemiology and drug safety. 2012.
100. Selim AJ, Berlowitz DR, Fincke G, Rosen AK, Ren XS, Christiansen CL, et al. Risk-adjusted mortality rates as a potential outcome indicator for outpatient quality assessments. Med Care. 2002; **40**(3): 237-45.
101. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. Am J Epidemiol. 2006; **163**(12): 1149-56.
102. Patrick AR, Schneeweiss S, Brookhart MA, Glynn RJ, Rothman KJ, Avorn J, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. Pharmacoepidemiol Drug Saf. 2011; **20**(6): 551-9.
103. Patorno E, Glynn RJ, Hernandez-Diaz S, Liu J, Schneeweiss S. Studies with many covariates and few outcomes: selecting covariates and implementing propensity-score-based confounding adjustments. Epidemiology. 2014; **25**(2): 268-78.
104. Muller-Oerlinghausen B, Wolf T, Ahrens B, Glaenz T, Schou M, Grof E, et al. Mortality of patients who dropped out from regular lithium prophylaxis: a collaborative study by the International Group for the Study of Lithium-treated patients (IGSLI). Acta Psychiatr Scand. 1996; **94**(5): 344-7.
105. Bocchetta A. Mortality follow-up of patients since commencing lithium therapy. J Clin Psychopharmacol. 2005; **25**(2): 197-9.
106. Christodoulou GN, Lykouras EP. Abrupt lithium discontinuation in manic-depressive patients. Acta Psychiatr Scand. 1982; **65**(5): 310-4.
107. Suppes T, Baldessarini RJ, Faedda GL, Tohen M. Risk of recurrence following discontinuation of lithium treatment in bipolar disorder. Arch Gen Psychiatry. 1991; **48**(12): 1082-8.
108. Scott J, Pope M. Nonadherence with mood stabilizers: prevalence and predictors. J Clin Psychiatry. 2002; **63**(5): 384-90.
109. Svarstad BL, Shireman TI, Sweeney JK. Using drug claims data to assess the relationship of medication adherence with hospitalization and costs. Psychiatr Serv. 2001; **52**(6): 805-11.
110. Schumann C, Lenz G, Berghofer A, Muller-Oerlinghausen B. Non-adherence with long-term prophylaxis: a 6-year naturalistic follow-up study of affectively ill patients. Psychiatry Res. 1999; **89**(3): 247-57.

111. Maarbjerg K, Aagaard J, Vestergaard P. Adherence to lithium prophylaxis: I. Clinical predictors and patient's reasons for nonadherence. Pharmacopsychiatry. 1988; **21**(3): 121-5.
112. Aagaard J, Vestergaard P. Predictors of outcome in prophylactic lithium treatment: a 2-year prospective study. J Affect Disord. 1990; **18**(4): 259-66.
113. Johnson RE, McFarland BH. Lithium use and discontinuation in a health maintenance organization. Am J Psychiatry. 1996; **153**(8): 993-1000.
114. Kessing LV, Sondergard L, Kvist K, Andersen PK. Adherence to lithium in naturalistic settings: results from a nationwide pharmacoepidemiological study. Bipolar Disord. 2007; **9**(7): 730-6.
115. Ray WA. Observational studies of drugs and mortality. N Engl J Med. 2005; **353**(22): 2319-21.
116. Lolich M, Vazquez GH, Alvarez LM, Tamayo JM. Psychosocial interventions in bipolar disorder: a review. Actas espanolas de psiquiatria. 2012; **40**(2): 84-92.
117. Colom F, Vieta E, Sanchez-Moreno J, Palomino-Otiniano R, Reinares M, Goikolea JM, et al. Group psychoeducation for stabilised bipolar disorders: 5-year outcome of a randomised clinical trial. The British journal of psychiatry : the journal of mental science. 2009; **194**(3): 260-5.
118. Kessing LV, Hansen HV, Hvenegaard A, Christensen EM, Dam H, Gluud C, et al. Treatment in a specialised out-patient mood disorder clinic v. standard out-patient treatment in the early course of bipolar disorder: randomised clinical trial. The British journal of psychiatry : the journal of mental science. 2013; **202**(3): 212-9.
119. Kessing LV, Hellmund G, Geddes JR, Goodwin GM, Andersen PK. Valproate v. lithium in the treatment of bipolar disorder in clinical practice: observational nationwide register-based cohort study. The British journal of psychiatry : the journal of mental science. 2011; **199**(1): 57-63.
120. Gallo JJ, Morales KH, Bogner HR, Raue PJ, Zee J, Bruce ML, et al. Long term effect of depression care management on mortality in older adults: follow-up of cluster randomized clinical trial in primary care. BMJ. 2013; **346**: f2570.
121. Schneider LS, Dagerman KS, Insel P. Risk of death with atypical antipsychotic drug treatment for dementia: meta-analysis of randomized placebo-controlled trials. JAMA : the journal of the American Medical Association. 2005; **294**(15): 1934-43.
122. Valenstein M, Kim HM, Ganoczy D, McCarthy JF, Zivin K, Austin KL, et al. Higher-risk periods for suicide among VA patients receiving depression treatment: Prioritizing suicide prevention efforts. J Affect Disord. 2009; **112**(1-3): 50-8.
123. Mander AJ, Loudon JB. Rapid recurrence of mania following abrupt discontinuation of lithium. Lancet. 1988; **2**(8601): 15-7.
124. Goodwin GM. Recurrence of mania after lithium withdrawal. Implications for the use of lithium in the treatment of bipolar affective disorder. Br J Psychiatry. 1994; **164**(2): 149-52.
125. Huybrechts KF, Brookhart MA, Rothman KJ, Silliman RA, Gerhard T, Crystal S, et al. Comparison of different approaches to confounding adjustment in a study on the

association of antipsychotic medication with mortality in older nursing home patients. Am J Epidemiol. 2011; **174**(9): 1089-99.
126. Garbe E, Kloss S, Suling M, Pigeot I, Schneeweiss S. High-dimensional versus conventional propensity scores in a comparative effectiveness study of coxibs and reduced upper gastrointestinal complications. Eur J Clin Pharmacol. 2013; **69**(3): 549-57.
127. Toh S, Garcia Rodriguez LA, Hernan MA. Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records. Pharmacoepidemiol Drug Saf. 2011; **20**(8): 849-57.
128. Le HV, Poole C, Brookhart MA, Schoenbach VJ, Beach KJ, Layton JB, et al. Effects of aggregation of drug and diagnostic codes on the performance of the high-dimensional propensity score algorithm: an empirical example. BMC medical research methodology. 2013; **13**: 142.
129. Brookhart MA, Rassen JA, Wang PS, Dormuth C, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? Med Care. 2007; **45**(10 Supl 2): S116-22.
130. Shack LG, Rachet B, Williams EM, Northover JM, Coleman MP. Does the timing of comorbidity affect colorectal cancer survival? A population based study. Postgrad Med J. 2010; **86**(1012): 73-8.
131. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005; **43**(11): 1130-9.
132. Kapur N, Cooper J, King-Hele S, Webb R, Lawlor M, Rodway C, et al. The repetition of suicidal behavior: a multicenter cohort study. Journal of Clinical Psychiatry. 2006; **67**(10): 1599-609.
133. Kim HM, Smith EG, Stano CM, Ganoczy D, Zivin K, Walters H, et al. Validation of key behaviourally based mental health diagnoses in administrative data: suicide attempt, alcohol abuse, illicit drug abuse and tobacco use. BMC Health Serv Res. 2012; **12**: 18.
134. Centers for Disease Control. National Center for Injury Prevention and Control. WISQARS Injury MOrtality Reports, 1999-2007. [cited June 2, 2012.]; Available from: http://webappa.cdc.gov/sasweb/ncipc/mortrate10_sy.html
135. Parikh NI, Gona P, Larson MG, Wang TJ, Newton-Cheh C, Levy D, et al. Plasma renin and risk of cardiovascular disease and mortality: the Framingham Heart Study. Eur Heart J. 2007; **28**(21): 2644-52.
136. Yerevanian BI, Koek RJ, Feusner JD. Pharmacotherapy and risk of suicidal behaviors among patients with bipolar disorder. JAMA. 2004; **291**(8): 939; author reply 40.
137. Scott J, Pope M. Self-reported adherence to treatment with mood stabilizers, plasma levels, and psychiatric hospitalization. Am J Psychiatry. 2002; **159**(11): 1927-9.
138. Licht RW, Vestergaard P, Rasmussen NA, Jepsen K, Brodersen A, Hansen PE. A lithium clinic for bipolar patients: 2-year outcome of the first 148 patients. Acta Psychiatr Scand. 2001; **104**(5): 387-90.

139. Manwani SG, Szilagyi KA, Zablotsky B, Hennen J, Griffin ML, Weiss RD. Adherence to pharmacotherapy in bipolar disorder patients with and without co-occurring substance use disorders. J Clin Psychiatry. 2007; **68**(8): 1172-6.
140. Smith EG, Kim HM, Ganoczy D, Stano C, Pfeiffer PN, Valenstein M. Suicide risk assessment received prior to suicide death by Veterans Health Administration patients with a history of depression. The Journal of clinical psychiatry. 2013; **74**(3): 226-32.
141. Kim HM, Smith EG, Ganoczy D, Walters H, Stano CM, Ilgen MA, et al. Predictors of suicide in patient charts among patients with depression in the Veterans Health Administration health system: importance of prescription drug and alcohol abuse. The Journal of clinical psychiatry. 2012; **73**(10): e1269-75.
142. Stone M, Laughren T, Jones ML, Levenson M, Holland PC, Hughes A, et al. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. BMJ. 2009; **339**: b2880.
143. Patrick AR, Schneeweiss S, Brookhart MA, Glynn RJ, Rothman KJ, Avorn J, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. Pharmacoepidemiology and Drug Safety. 2011; **20**(6): 551-9.
144. Oquendo MA, Stanley B, Ellis SP, Mann JJ. Protection of human subjects in intervention research for suicidal behavior. Am J Psychiatry. 2004; **161**(9): 1558-63.
145. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA. 2007; **297**(3): 278-85.
146. Eng PM, Seeger JD, Loughlin J, Clifford CR, Mentor S, Walker AM. Supplementary data collection with case-cohort analysis to address potential confounding in a cohort study of thromboembolism in oral contraceptive initiators matched on claims-based propensity scores. Pharmacoepidemiol Drug Saf. 2008; **17**(3): 297-305.