APPROPRIATE USE AND VALUE OF SURVEILLANCE AMONG MEDICARE PATIENTS WITH NON-MUSCLE-INVASIVE BLADDER CANCER

Mihaela V. Georgieva

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Health Policy and Management in the Gillings School of Global Public Health.

Chapel Hill
2018

Approved by:

Stephanie B. Wheeler

Matthew E. Nielsen

Justin G. Trogdon

Michaela A. Dinan

Morris Weinberger

ABSTRACT

Mihaela V. Georgieva: Appropriate Use and Value of Surveillance among Medicare Patients with Non-Muscle-Invasive Bladder Cancer
(Under the direction of Stephanie B. Wheeler)

Bladder cancer patients have the highest median age at diagnosis of 73 years compared with all other cancer types and often live with substantial disease and comorbidity burden. Due to high recurrence rates, intensive surveillance strategies, and expensive therapies, bladder cancer has the highest lifetime costs per patient from diagnosis to death. Regular surveillance cystoscopy is recommended for patients with non-muscle invasive bladder cancer (NMIBC) of all ages to detect potential recurrences, despite a lack of large randomized controlled trials examining how the use of cystoscopy affects patient outcomes.

The overall objectives of this dissertation were (1) to investigate factors associated with receipt of surveillance cystoscopy; (2) to characterize survival outcomes of NMIBC patients undergoing surveillance; and (3) to examine the cost-effectiveness of three different risk-stratified and uniform surveillance recommendations. We used the Surveillance Epidemiology and End Results (SEER)-Medicare data from 2000 to 2014 to assess disease characteristics and outcomes. We also developed a patient-level simulation model to quantify the health-economic impact of different frequencies of surveillance over five years. In Aim 1, we found that NMIBC patients aged ≥85 years, those with poor disability status, and those having ≥3 comorbidities at diagnosis were least likely to undergo recommended (≥7 cystoscopies) or low-intensity (≥4 cystoscopies) surveillance over the first two years post-diagnosis. As the age at diagnosis and the

number of comorbid conditions increased, the odds of receiving recommended cystoscopy frequency as well as the rate of cystoscopy decreased. In Aim 2, older patients (≥75 vs. 66-74 years) and those with poor disability status at diagnosis had higher cumulative incidence of both bladder-cancer and other-cause death, regardless of frequency of cystoscopy. In Aim 3, low-intensity risk-stratified surveillance, with cystoscopy frequency increasing progressively with risk, was associated with different trade-offs such as lower costs and fewer false positive cases per patient, compared with a more frequent high-intensity risk-stratified approach and uniform surveillance.

This research highlights the importance of age, comorbidities, functional status, and risk-stratification on receipt of surveillance and outcomes of NMIBC patients. Additionally, findings from our study suggest intermediate-risk patients may benefit from less frequent surveillance than high-risk patients.

ACKNOWLEDGEMENTS

They say it takes a village to raise a child and I'd say the same about completing a dissertation. This research and milestone would not have been possible without the unwavering support of my mentors, family, and friends. I thank my advisor and dissertation chair, Stephanie Wheeler, for her direction and sage advice throughout the last four years and especially during the dissertation phase. Stephanie encouraged me to pursue the tough but important questions and challenge myself with new methods beyond what we learned in the classroom.

She also helped me form the dissertation "dream team" who guided and supported me every step of the way: Matthew Nielsen, Justin Trogdon, Jennifer Lund, Michaela Dinan, and Morris Weinberger. This dissertation research was largely inspired by Matt Nielsen. From our conversation about bladder cancer at Friends' Café during my first year in the program to many drafts and analyses later, Matt's perspective and insights have been invaluable. I am really grateful to Justin Trogdon and Jenny Lund for helping me navigate different methods and sharing many resources with me – you taught me a lot! I thank my DCRI mentor, Michaela Dinan, for introducing me to the intricacies of using SEER-Medicare data, bringing me on a project through which I could access the data for my dissertation, and providing me with leadership opportunities. To Morris Weinberger – thank you for all the support since the first call, notifying me of my admittance to the program.

I would also like to thank my collaborators, faculty, and staff at the DCRI: Charles Scales and Shelby Reed—for the research opportunities; Lauren Wilson and Yanhong Li—for answering my programming questions and helping me troubleshoot cryptic SAS error messages.

I thank Amy Davidoff for allowing me to use the disability status measure developed by her group and sharing the macro with me.

I am thankful for the funding I have received to support this dissertation research and my graduate studies, including the DCRI Pre-doctoral Fellowship, the Executive Director's Pathway Fellows Supplemental Funding Award, Wellesley College Graduate Horton-Hallowell Fellowship, and the University Cancer Research Fund Award.

Last, but not least, I am forever indebted to my family and friends for their support and encouragement. To my mom and dad, thank you for always believing in me, encouraging me to follow my dreams, even if they took me thousands of miles away from you, and instilling in me the love of learning. To my grandparents and sister – you have always been a source of inspiration. To my family away from home, grad school and college friends – thank you for making this journey lighter and more fun.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AJCC	American Joint Committee on Cancer
ALS	Amyotrophic lateral sclerosis
ASCOG	American College of Surgeons Oncology Group
AUA	American Urological Association
BCG	Bacillus Calmette-Guérin
CCI	Charlson Comorbidity Index
CCW	Chronic Conditions Warehouse
CDC	Centers for Disease Control and Prevention
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CIS	Carcinoma in situ
CMS	Centers for Medicare and Medicaid Services
CPT	Current Procedural Terminology
CSS	Cancer-specific survival
DCRI	Duke Clinical Research Institute
DME	Durable medical equipment
DSA	Deterministic sensitivity analysis
DUA	Data Use Agreement
EAU	European Association of Urology
ECOG	Eastern Cooperative Oncology Group
ED	Emergency department
EDB	Enrollment Database

EEE	Extended estimating equations
EORTC	European Organization for Research and Treatment of Cancer
ESRD	End-stage renal disease
FP	False positive
GEE	Generalized estimating equations
GLM	Generalized linear model
HCPCS	Healthcare Common Procedure Coding System
ICD-9	International Classification of Diseases 9th edition
ICER	Incremental cost-effectiveness ratio
IRB	Institutional Review Board
LVI	Lymphovascular invasion
MBSF	Master Beneficiary Summary File
MedPAR	Medicare Provider Analysis and Review
MMC	Mitomycin C
NCI	National Cancer Institute
NMIBC	Non-muscle invasive bladder cancer
NOS	Not Otherwise Specified
NPI	National Provider Identifier
PEDSF	Patient Entitlement and Diagnosis Summary File
PS	Performance status
PSA	Probabilistic sensitivity analysis
PUNLMP	Papillary urothelial neoplasm of low malignant potential
QALY	Quality-adjusted life year

SEER	Surveillance Epidemiology and End Results
SSA	Social Security Administration
SUO	Society of Urologic Oncology
TNM	Tumor, Node, Metastasis
TURBT	Transurethral resection of the bladder tumor
UPIN	Unique Physician Identification Number
US	United States

CHAPTER 1: INTRODUCTION

Overview

The rising costs of cancer care present a substantial burden for the healthcare system and most importantly, for patients. Additionally, the aging of the population, the increased incidence/prevalence of cancer, and the increased complexity of treatments present further challenges to the delivery of high-quality cancer care and call for evidence-based care and scientific research to inform medical decisions (Institute of Medicine 2013). High-cost care does not necessarily translate into high-quality care or improved outcomes, as shown by the *Choosing Wisely* initiative (Howard and Gross 2015). The recent emergence of different value frameworks has been one effort to quantify the health benefits of cancer treatments relative to their cost and help promote shared medical decision making (Chandra, Shafrin, and Dhawan 2016). Determining the appropriate intensity of care, particularly with respect to cancer surveillance, also requires quantifying the long-term benefits and costs associated with different courses of action. Surveillance approaches following active treatment for cancer have been identified as the highest priority topic for cancer-related comparative effectiveness research (Greenberg et al. 2013).

One area characterized by considerable variation in practice and uncertainty regarding optimal surveillance choices is that of early stage bladder cancer. Recommendations from existing clinical guidelines for the management of non-muscle-invasive bladder cancer (NMIBC) are largely consensus-based and vary across different professional societies and countries, with approaches ranging from intensive one-size-fits-all surveillance schedules, historically in the United States, to patient-level risk-stratification, historically in Europe (Chang et al. 2016; Hall

et al. 2007; Power and Izawa 2016; Babjuk et al. 2016; Babjuk et al. 2013; Babjuk et al. 2011). Prior studies have found lack of adherence to bladder cancer surveillance guidelines and underutilization of care for patients with NMIBC in the United States arguably due to unnecessarily intensive recommendations for use of surveillance cystoscopy (Schrag et al. 2003; Chamie et al. 2011; Chamie et al. 2012). In June 2016, the American Urological Association (AUA) and Society of Urologic Oncology (SUO) released updated guidelines for diagnosis and treatment of NMIBC recommending a risk-stratified approach to surveillance cystoscopy based on known risks for recurrence and progression (Chang et al. 2016). However, surprisingly, little is known about how different frequencies of surveillance affect patient outcomes and costs. Moreover, non-disease related factors such as frailty, functional status, and comorbid conditions could impact both receipt of surveillance and outcomes (Taylor and Kuchel 2009; Soria et al. 2016). Given the high costs associated with bladder cancer treatment and increased disease burden as the US population ages, it is important to understand what the implications of surveillance frequency and risk-stratification are for providing high-quality, high-value care for older NMIBC patients. We used the generally accepted definition of value as a measure of outcomes achieved per unit of monetary expenditure (Feeley et al. 2010).

Purpose of the Dissertation

The long-term goal of this research is to improve the efficiency, quality, and value of cancer care by limiting unnecessary and/or potentially harmful interventions and promoting effective and patient-centered cancer treatment and surveillance. The overall objective of this dissertation was to examine real-world surveillance patterns and quantify the health-economic impact of surveillance among older patients with NMIBC. The central hypothesis was that less

frequent NMIBC surveillance accounting for patient-specific characteristics and heterogeneity would result in similar patient outcomes, compared to the historic US guidelines recommending uniform surveillance cystoscopy every three months, and would be associated with lower costs. The specific aims were:

1. Examine surveillance patterns and factors associated with receipt of recommended surveillance cystoscopy for NMIBC patients.
    a. Characterize differences between patients receiving recommended (≥7 cystoscopies) or at least low-intensity (≥4 cystoscopies) surveillance over the first two years after diagnosis.
    b. Examine the effect of comorbidities, age, and disability/functional status on receipt of recommended and low-intensity surveillance as well as on the number of cystoscopies received during the first two years after diagnosis, after controlling for other demographic and clinical characteristics.
2. Examine the association between surveillance cystoscopy use and NMIBC survival within the Medicare population.
    a. Characterize survival patterns and assess all-cause, bladder-cancer-specific, and other-cause mortality among NMIBC patients after receipt of recommended (≥7 vs. <7 cystoscopies) and low-intensity (≥4 vs. <4 cystoscopies) during the first two years after diagnosis.
    b. Examine differences in bladder-cancer-specific and other-cause mortality by patients' age at diagnosis, comorbidities, and disability status.

3. Assess the cost-effectiveness of the US risk-stratified surveillance guidelines compared with the historic recommendations of uniform surveillance among Medicare patients with NMIBC.
    a. Quantify the health-economic outcome in terms of total quality-adjusted life years (QALY) gained, total costs, total recurrent cases and progressed cases detected, total deaths, and total false positive (FP) cases under the historic uniform recommendations, a low intensity risk-stratified approach and a high intensity risk-stratified approach, based on the 2016 American Urological Association (AUA)/ Society of Urologic Oncology (SUO) guidelines.
    b. Compare the incremental cost-effectiveness ratios (ICERs) of cost per additional QALY gained, cost per death averted, cost per recurrence detected, cost per progression averted, and cost per FP averted across the three strategies.

Covariates used in the Aims 1 and 2 analyses included patient age at diagnosis, race/ethnicity, marital status at diagnosis, tumor grade, T classification (Ta, Tis, T1) based on the American Joint Committee on Cancer (AJCC) TNM staging system (Edge et al. 2010), prior history of cancer, year of diagnosis, history of state buy-in in the year prior diagnosis (proxy for whether a beneficiary was enrolled in a state-administered Medicaid program), SEER region, and extent of urbanization at patients' residence. We assessed functional status in the 12 months prior to diagnosis using the Disability Status (DS) measure (Davidoff et al. 2013; Davidoff 2014). Comorbidities were measured using the Klabunde adaptation of the Charlson comorbidity index modified for cancer patients (Klabunde et al. 2007; Klabunde et al. 2000). We assessed patients'

zip code-level socioeconomic information by using US Census data to derive quartiles of median household income in subject's zip code and proportion of adult residents with less than high school education in subject's zip code. Surgeon volume was defined as the number of transurethral resections performed by each patients' treating physician in the year prior to patient's diagnosis.

Aim 1 was analyzed using multilevel logistic and Poisson regression to examine frequency of surveillance, defined as ≥7 (recommended), ≥4 cystoscopies (low-intensity), and number of cystoscopies during the first two years post-diagnosis. Aim 2 was analyzed using propensity-score weighted Fine-Gray competing risk regression and Cox proportional hazards model to assess all-cause, bladder-cancer-specific and other-cause mortality after receipt of high-intensity (≥7 vs. <7 cystoscopies) or low-intensity surveillance (≥4 vs. <4 cystoscopies) during the first two years after diagnosis. Aim 3 was analyzed using a patient-level simulation model and risk-stratification to compare the three different surveillance approaches.

Significance and Contribution

This research is innovative in several major ways. It is the first study to evaluate the functional status of NMIBC patients at diagnosis, using as a proxy the disability status measure, a novel claims-based prediction algorithm for performance status in older adults (Davidoff et al. 2013). We examined patients' disability status at diagnosis as a predictor of surveillance use and incorporated it in the survival models to alleviate treatment selection bias and confounding by frailty. Furthermore, no studies have examined recent trends in NMIBC surveillance in a broader cohort of NMIBC patients with both high- and low-grade tumors. Lastly, this is the first study to investigate the cost-effectiveness of the 2016 AUA/SUO guidelines using risk-stratification,

compared with the historic US surveillance approach, recommending uniform cystoscopic evaluation for all patients (Chang et al. 2016; Hall et al. 2007).

The research presented in this dissertation helps to inform the decision making of providers, insurers, and patients and quantifies the value and risk-benefit trade-offs of surveillance for older NMIBC patients. The analyses also help improve the understanding of age-related factors associated with receipt of surveillance and identify patient groups that may benefit from further interventions aimed at promoting appropriate frequency of surveillance. Findings from this work may also provide insights in improving surveillance approaches following active treatment for cancer among other cancer types and in particular, among older cancer survivors with multiple chronic conditions.

Structure of the Dissertation

Sections of the dissertation are organized as follows: Chapter 2 presents a detailed review of NMIBC disease characteristics, summary of existing domestic and international clinical practice guidelines for NMIBC surveillance, as well as current risk-stratification algorithms and their use in different guidelines. We also discuss the current literature on bladder cancer surveillance, adherence to surveillance recommendations, and limitations of the existing observational studies. The purpose of Chapter 2 is to provide background and motivation for the dissertation study.

Chapter 3 provides an overview of the methods used, including the conceptual framework motivating this research, data sources and study population, variable definitions and measurement, study hypotheses, as well as detailed analytical approaches by aim, and description of sensitivity analyses and expected limitations. Chapters 4 through 6 are individual manuscripts corresponding to aims 1-3, respectively, and are intended for independent

submission for peer-reviewed publication. Finally, Chapter 7 summarizes the strengths and limitations of this work, policy implications, and future research directions. An appendix with supplementary tables and materials, referenced in earlier chapters, and a full bibliography are provided at the end of the dissertation.

CHAPTER 2: LITERATURE REVIEW

Disease Characteristics

Bladder cancer is the sixth most common cancer in the United States, and its incidence has increased from approximately 50,000 new cases diagnosed in 1990 to more than 80,000 new cases expected in 2018 (Silverberg, Boring, and Squires 1990; Siegel, Miller, and Jemal 2018). Over the same time period, the median age at diagnosis has increased from 65 to 73 years, the highest age among all cancer sites (American Cancer Society 2016). Most incident cases (70%) are non-muscle-invasive bladder cancer (NMIBC) at the time of presentation but these patients have highly variable risks of recurrence of NMIBC and progression to the potentially lethal phenotype of muscle-invasive (≥T2) disease (Burger et al. 2013; Sylvester et al. 2006). Low-grade NMIBC constitutes about two-thirds of all NMIBC cases; in this group approximately 50% recur, and <5% progress to muscle-invasive disease (Sylvester et al. 2006). In the United States, the incidence rate of non-invasive papillary (Ta) predominantly low-grade bladder cancer was found to increase significantly between 1988-2006, while over the same time period the rates of carcinoma in situ (Tis) and lamina propria-invasive (T1) disease decreased (Nielsen et al. 2014). These trends were most pronounced among older patients aged 65 years and above, suggesting changing disease characteristics among older cohorts (Nielsen et al. 2014).

Long-term follow-up has shown that the progression rate varies by tumor grade and T stage and is as low as 6% for low-grade Ta and as high as 17% for high-grade T1 (Palou et al. 2012; Leblanc et al. 1999). The survival prognosis for patients with NMIBC is relatively favorable, with cancer-specific survival at 5 years ranging from 98.5% in patients with low-grade

Ta to 88.7% in patients with high-grade T1 disease (Cambier et al. 2016). However, cancer-specific survival after progression from high-risk (e.g., high-grade T1) NMIBC to MIBC have been found to be much lower: 32% (range: 13%-64%) in 7 prospective trials with a median follow-up of 52-123 months and 37% (range: 7%-59%) in 12 retrospective studies with a median follow-up of 48-107 months (Van Den Bosch and Witjes 2011).

Comorbidities and comorbidity-associated events also represent very important causes of mortality in bladder cancer patients, who have one of the highest comorbidity burdens compared with other cancer patients, increasing progressively with age (Bluethmann, Mariotto, and Rowland 2016). Among patients newly diagnosed with bladder cancer, the 5-year cancer-specific mortality rate was found to vary between 1% and 59%, and other-cause mortality rate between 6% and 90%, depending on the tumor type and patient age (Noon et al. 2013). Even among high-risk bladder cancer patients treated with radical cystectomy, between 8.5% and 27.1% of deaths were attributable to other-cause mortality at 5 years after cystectomy (Lughezzani et al. 2011).

Despite the variable but generally low progression rates, NMIBC has very high rates of recurrent NMIBC (Sylvester et al. 2006). Given the high recurrence rates, regular surveillance cystoscopy is recommended for patients with NMIBC of all ages to detect potential recurrences or progression to MIBC. The evidence of a growing subgroup of predominantly older patients with lower-risk disease and high comorbidity burden warrants further scrutiny of clinical practice recommendations for NMIBC, which currently do not explicitly consider age, comorbidities, or functional impairment when determining appropriate intensity of care.

Overview of Different Surveillance Recommendations

For patients with early stage NMIBC, surveillance is based on repetitive invasive endoscopy procedures, often for the remainder of a patient's lifespan. Historically, guidelines in the United States recommended surveillance cystoscopy every 3 months for 2 years, every 6 months for 2 years, and annually thereafter, even though neither the ideal interval, nor the duration of follow-up cystoscopy were defined (Hall et al. 2007). The 2016 AUA/SUO guidelines (Chang et al. 2016) recommended the following risk-stratified approach to surveillance:

1) "For a low-risk patient whose first surveillance cystoscopy is negative for tumor, a clinician should perform subsequent surveillance cystoscopy six to nine months later, and then annually thereafter; surveillance after five years in the absence of recurrence should be based on shared-decision making between the patient and clinician. (Moderate Recommendation; Evidence Strength: Grade C)"
2) "For an intermediate-risk patient whose first surveillance cystoscopy is negative for tumor, a clinician should perform subsequent cystoscopy with cytology every 3-6 months for 2 years, then 6-12 months for years 3 and 4, and then annually thereafter. (Expert Opinion)"
3) "For a high-risk patient whose first surveillance cystoscopy is negative for tumor, a clinician should perform subsequent cystoscopy with cytology every three to four months for two years, then six months for years three and four, and then annually thereafter. (Expert Opinion)"

Given the lack of prospective studies that compare outcomes among differing cystoscopic surveillance regimens for intermediate- and high-risk NMIBC patients (unlike the case for low-

risk patients), the Panel recommendations for frequency of surveillance for the intermediate- and high-risk groups were based on consensus and historic precedent (Chang et al. 2016). The Panel highlighted the "urgent need for studies to determine if less stringent follow up regimens can be employed without significantly affecting oncologic outcomes in these [intermediate-risk and high-risk] patients."

International guidelines using risk-stratified approaches to surveillance cystoscopy have already adopted less stringent surveillance schedules (Table **1**). The European Association of Urology (EAU) guidelines recommend cystoscopy for low-risk patients at three months, nine months, and annually thereafter; for intermediate-risk patients: at three months, every three to six months for five years, and annually thereafter; for high-risk patients: every three months for two years, every six months for three years, and annually thereafter (Babjuk et al. 2016; Babjuk et al. 2013; Babjuk et al. 2011). The National Institute for Health and Care Excellence (NICE) guideline for bladder cancer diagnosis and management in the United Kingdom recommend cystoscopic follow-up for low-risk patients at three and 12 months after diagnosis and discharge to primary care for patients with no recurrence of the bladder cancer within 12 months. For intermediate-risk patients, NICE recommends cystoscopy at three months, nine months, 18 months, and annually thereafter for five years; for high risk patients: every three months for two years, every six months for the next two years, and annually thereafter (National Collaborating Centre for Cancer 2015).

Table 1. Comparison of existing guidelines for the diagnosis and management of NMIBC

	US Historic Recs	**AUA/SUO Guidelines (2016)**			**EAU Guidelines (2008-2016)**			**NICE Guidelines (2015)**		
	All	Low Risk	Int Risk	High Risk	Low Risk	Int Risk	High Risk	Low Risk	Int Risk	High Risk
Cystoscopy	3 months/ 2 years	3-4 months	3 months	3-4 months /2 years	3 months	3 months	3 months / 2 years	3 months	3 months	3 months / 2 years
	6 months/ 2-3 years	6-9 months	3-6 months / 2 years	6 months /2 years	9 months	3-6 months/ 5 years	6 months / 3 years	12 months	9 months	6 months /2 years
	once/ year	once/ year	6-12 months /2 years	once/ year	once/ year	once/ year	once/ year		18 months	once/ year
			once/ year						once/ year	
Total follow-up	N/A	5 years	N/A	N/A	5 years	N/A	N/A	12 months	5 years	N/A

Abbreviations: US, United States; recs, recommendations; AUA, American Urological Association; SUO, Society of Urologic Oncology; EAU, European Association of Urology; NICE, National Institute for Health and Care Excellence; Int, intermediate; N/A, not available.

Risk-stratification

Prognostic markers in NMIBC that help assess the risk of recurrence and progression to muscle-invasive bladder cancer (MIBC) include clinical factors such as history of recurrence and pathological factors such as stage, grade, cancer size, presence of carcinoma in situ (CIS), number of cancers, variant pathology, lymphovascular invasion (Chang et al. 2016; Sylvester et al. 2006). The European Organization for Research and Treatment of Cancer (EORTC) risk tables used six clinical and pathological factors (number of tumors, tumor size, prior recurrence rate, T category, CIS, and grade) to stratify patients into risk groups and derive annual probabilities of recurrence and progression (Sylvester et al. 2006). The EORTC risk tables (Sylvester et al. 2006) have been validated in several studies which demonstrate that the tables successfully stratify patients into risk groups for recurrence and progression, although they tend to overestimate the risk of recurrence in all risk groups and the risk of progression in high risk

groups (Seo et al. 2010; Fernandez-Gomez et al. 2011; Altieri et al. 2012; T. Xu et al. 2013; Lammers et al. 2014; Hernandez et al. 2011).

There is no universally accepted classification of risk in NMIBC. Summary of the risk-stratification classifications proposed by the AUA/SUO, EAU, and NICE is presented in Table **2**, Table **3**, and Table **4**. One of the main differences in the risk-stratification algorithms is that unlike the EAU and NICE definitions, the AUA/SUO classification also includes any BCG failure in high grade cases, any lymphovascular invasion (LVI), and any high grade prostatic urethral involvement among the criteria for high-risk NMIBC. The NICE guidelines did not include presence of LVI as a risk factor due to their assessment of low quality evidence suggesting that the presence of LVI increases the risk of recurrence, progression, and disease-specific survival in small study samples (National Collaborating Centre for Cancer 2015; Lotan et al. 2005; Kikuchi et al. 2009). While both the AUA/SUO and NICE classifications include variant histology as a high-risk factor, the NICE definition is more restrictive, including only aggressive variants of urothelial carcinoma (e.g., micropapillary or nested variants).

Table 2. 2016 AUA/SUO risk stratification algorithm for NMIBC

Low Risk	**Intermediate Risk**	**High Risk**
Low grade solitary Ta ≤3 cm Papillary urothelial neoplasm of low malignant potential	Recurrence within 1 year, low grade Ta Solitary low grade Ta >3 cm Low grade Ta, multifocal High grade Ta, ≤3 cm Low grade T1	High grade T1 Any recurrent, high grade Ta High grade Ta, >3 cm (or multifocal) Any CIS Any BCG failure in high grade case Any variant histology Any lymphovascular invasion Any high grade prostatic urethral involvement

Source: Chang et al. 2016

Table 3. EAU risk stratification algorithm for NMIBC

Low Risk	Intermediate Risk	High Risk
Primary, solitary, Ta, LG (low grade)/G1, < 3 cm, no CIS	All tumors not defined in the two adjacent categories (between the category of low and high risk)	Any of the following: • T1 tumor • HG (high grade)/G3 tumor • CIS • Multiple and recurrent and large (>3 cm) • Ta G1G2 tumors (all conditions must be present in this point)

Source: Babjuk et al. 2011, 2013, 2016.

Table 4. NICE risk stratification algorithm for NMIBC

Low Risk	Intermediate Risk	High Risk
• Solitary pTaG1 < 3 cm • Solitary pTaG2 (low grade) <3 cm • Any PUNLMP (papillary urothelial neoplasm of low malignant potential)	• Solitary pTaG1 ≥3 cm • Multifocal pTaG1 • Solitary pTaG2 (low grade) ≥3 cm • Multifocal pTaG2 (low grade) • pTaG2 (high grade) • Any pTaG2 (grade not further specified) • Any low‐risk non‐muscle‐invasive bladder cancer recurring within 12 months of last tumor occurrence	Any of the following: • pTaG3 • pT1G2 • pT1G3 • pTis (CIS) • Aggressive variants of urothelial carcinoma, for example micropapillary or nested variants

Source: NICE Guideline for Bladder Cancer (2015)

Economic Burden of Surveillance

Due to high recurrence rates, intensive surveillance strategies, and expensive therapies, the economic burden of bladder cancer is substantial (Svatek et al. 2014; Noyes, Singer, and Messing 2008). In the United States, the annual national cost of bladder cancer care in 2010 was estimated to be $3.98 billion and is expected to rise to $5 billion by 2020 (Mariotto et al. 2011). Lifetime per capita costs of bladder cancer, estimated between $96,000 and $187,000 in 2001 US dollars (more than $260,000 in 2015 US dollars), is higher than the cost of any other cancer from diagnosis to death (Botteman et al. 2003; Crawford, Church, and Akin 2016). Given the well-established trends in long-term survival in the literature and expensive treatments, it is expected

that bladder cancer will continue to have the highest lifetime costs per patient from diagnosis to death, exceeding those of other common cancers, including colorectal, breast, prostate, and lung cancers (Yabroff et al. 2011; Yabroff et al. 2008; Yeung, Dinh, and Lee 2014). A retrospective study of 208 bladder cancer patients found that surveillance and the management of recurrences accounted for approximately 60% of the lifetime cost (Avritscher et al. 2006). If all patients diagnosed with NMIBC in the last year received recommended surveillance care with no recurrences (a perfect case scenario), the cost to the health care system for the management of this cohort for 5 years would exceed $150 million (Avritscher et al. 2006).

Adherence to Surveillance Recommendations in the United States

Prior analyses using the Surveillance, Epidemiology, and End Results (SEER) cancer registry data linked with Medicare claims have indicated that, in practice, adherence to historically recommended NMIBC surveillance protocols is low. Only 40% of patients with NMIBC diagnosed in SEER-Medicare between 1992 and 1996 underwent surveillance once every six months for three years (Schrag et al. 2003). Chamie et al. found even lower rates of adherence to historic surveillance cystoscopy recommendations: among high-grade NMIBC patients diagnosed between 1992-2002, only 4.9% received eight or more cystoscopies and 33.6% received at least four cystoscopies over the first two years after diagnosis (Chamie et al. 2011). Unexplained provider-level factors contributed significantly to the low compliance rate (Chamie et al. 2011). In a follow-up study by the same research group, compliance with ≥4 cystoscopies, ≥4 cytologies, and BCG instillation was found to be lower among patients aged ≥80 and higher among those with undifferentiated, Tis, and T1 tumors, and among patients diagnosed after 1997 (Chamie et al. 2012). Patients compliant with these measures had a lower hazard of bladder-cancer mortality (hazard ratio, 0.41; 95% confidence interval, 0.18-0.93) than

those who received < 4 cystoscopies, < 4 cytologies, and no BCG (Chamie et al. 2012). However, all previous studies on bladder cancer surveillance lacked important information about patients' functional status, which may influence patients' and providers' decisions to undertake regular surveillance cystoscopy as well as patient outcomes (Schrag et al. 2003; Chamie et al. 2011; Chamie et al. 2012; Hollingsworth et al. 2010).

Treatment intensity, defined as all Medicare expenditures per patient incurred within the first 2 years after bladder cancer diagnosis, was used as another measure of health care utilization among patients with early stage bladder cancer (Hollenbeck et al. 2009; Hollingsworth et al. 2010; Strope et al. 2010). Treatment intensity was driven largely by patient characteristics: 23% of the variation in initial treatment intensity was determined by patient-level factors vs. 9.2% determined by provider factors (Hollingsworth et al. 2010). However, relatively little of the variation was accounted for by commonly measured patient-level factors such as age, comorbidities, tumor grade, and stage, suggesting that other unmeasured patient-level factors such as patient preferences or functional status might be influencing the variability in treatment intensity. While treatment intensity was appropriately aligned with clinical characteristics such as age, comorbidity, tumor stage, and grade, it did not match the inferred disease risk for about half of the highest and lowest risk patients, suggesting that a considerable proportion of patients with early-stage bladder cancer might be receiving too much or too little care (Strope et al. 2010). However, the definitions of disease risk used by Strope and colleagues were not based on any formal risk-stratification algorithm, neither did they capture the risks of recurrence and progression. Instead, the authors identified groups of patients with very low risk and very high risk of bladder cancer-related death, thus potentially misclassifying patients.

There are several limitations of defining treatment intensity based on bladder cancer expenditures rather than on specific procedures and services used. Using only Medicare payments for all inpatient and outpatient claims associated with a bladder cancer diagnosis as a proxy for treatment intensity fails to account for costs associated with procedural complications or frequency and indication of services received. A patient undergoing a costly cystectomy, a radical surgery to remove the bladder, may be classified as receiving the same level of high treatment intensity as a patient receiving a high volume of less intensive routine care or regular surveillance. Similarly, it is problematic to investigate the association between cancer-specific survival and treatment intensity based on Medicare payments due to potential endogeneity. Healthier patients are likely to both live longer and have lower Medicare spending/treatment intensity within the first two years after diagnosis. Therefore, the findings that treatment intensity is not associated with improved cancer-specific survival (Hollenbeck et al. 2009; Hollingsworth et al. 2010) merely demonstrate a lack of correlation between Medicare payments for bladder cancer and survival rather than any plausible causal link. Detailed review of the discussed studies examining receipt of health care services among early stage bladder cancer patients in the United States is presented in Appendix 2.1. Review of prior studies examining receipt of surveillance among bladder cancer patients in the Unites States is presented in Appendix 2.2.

CHAPTER 3: RESEARCH DESIGN AND METHODS

Overview and Rationale

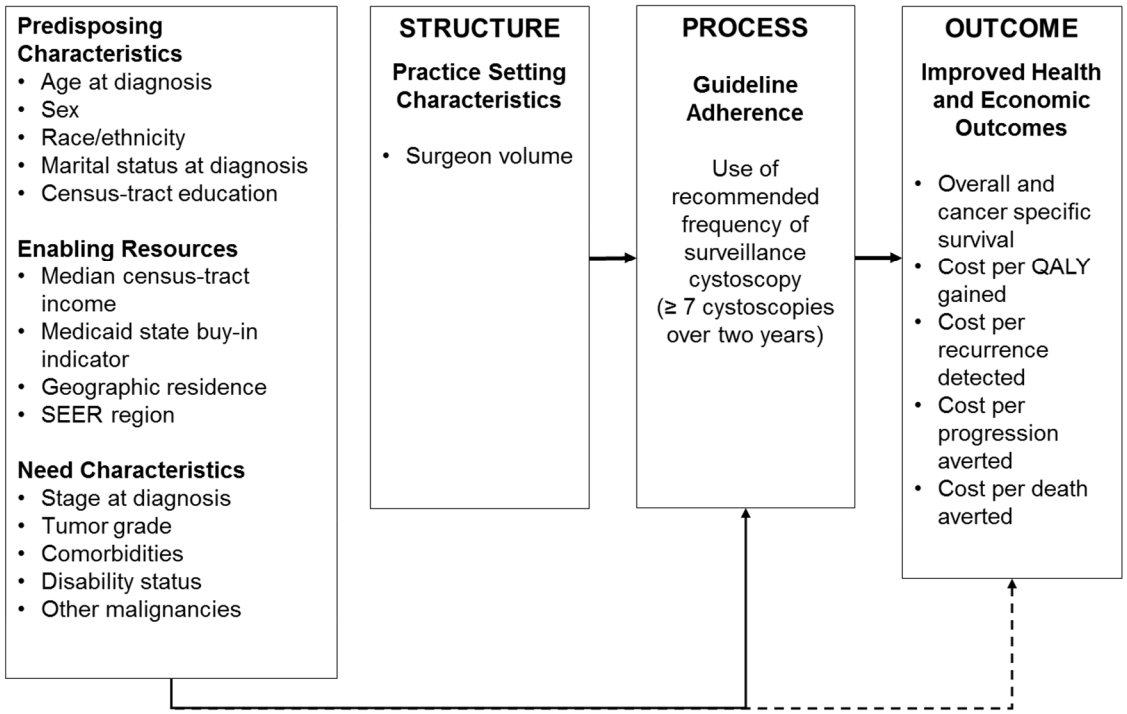
The goal of this dissertation research is to examine the value of surveillance for older patients with non-muscle-invasive bladder cancer (NMIBC) by assessing disease characteristics, surveillance patterns, and outcomes among Medicare beneficiaries with NMIBC. The overall study design for Aims 1 and 2 consisted of a retrospective cohort study of Medicare beneficiaries diagnosed with NMIBC from 2001 to 2012. Aim 3 used a patient-level simulation model calibrated with data from SEER-Medicare, clinical trials, and the published literature to compare the cost-effectiveness of the new 2016 AUA/SUO risk-stratified guidelines with the historic US recommendations of uniform surveillance for all patients, regardless of their risk of recurrence and progression.

Conceptual Model

This research is grounded in a conceptual framework informed by both Andersen's Behavioral Model of Health Care (Andersen 1995) and Donabedian's Quality of Medical Care framework (Donabedian 1966). Andersen's model suggests that use of health services is a function of 1) predisposing characteristics (e.g., demographic factors), 2) enabling resources which allow the use of services, and 3) need for those services (Andersen 1995). The framework proposed by Donabedian consists of three elements: structure, process, and outcomes (Donabedian 1966). The underlying assumption for this model is that good structure increases the likelihood of good process and that, in turn, increases the likelihood of a good outcome

(Donabedian 1966). The proposed conceptual model combining Andersen's and Donabedian's frameworks allows to evaluate the influence of both patient-level and practice setting characteristics on receipt of recommended surveillance as well as on patient outcomes and costs.

Figure 1. Conceptual framework



Note: Conceptual framework informed by both Andersen's Behavioral Model of Health Care (Andersen 1995) and Donabedian's Quality of Medical Care framework (Donabedian 1966).

Donabedian defines structure as the attributes of the setting in which care occurs (Donabedian 1966). Structural elements are attributes of the site of care, including characteristics of physicians and their practice settings, which may influence health outcomes (Brook, McGlynn, and Cleary 1996). Since surveillance cystoscopy is typically administered in an outpatient setting and does not require hospitalization, we were not able to examine the association between hospital characteristics and patient surveillance. However, the structural elements we examined were distance traveled to cystoscopy provider and surgeon case volume. Surgeon case volume was defined as a categorical variable measuring caseload for transurethral

resections for each surgeon, following prior work (Chamie et al. 2011; Chamie et al. 2012). Surgeons with higher caseloads for transurethral resections might also be more likely to use more frequent surveillance cystoscopy due to inherent preferences or proclivity for patient workup and endoscopic evaluation. Therefore, it is important to account for these characteristics in the multivariable models examining factors associated with receipt of surveillance cystoscopy.

Process encompasses the components of the encounter between provider and patients such as whether clinical practice guidelines are followed. Process was assessed by examining receipt of surveillance cystoscopy among the study population. Outcome is defined as the effects of care on the health status of patients and populations. We evaluated survival, costs, as well as cost-effectiveness measures such as additional cost per quality adjusted life years (QALYs) gained and additional cost per death averted associated with different surveillance strategies.

Examining predisposing, enabling, and need characteristics is key to understanding the association between patient-level factors and surveillance cystoscopy use. Predisposing characteristics included age at diagnosis, sex, race/ethnicity, marital status, and education and were obtained from the SEER Patient Entitlement and Diagnosis Summary File (PEDSF). Enabling resources are those that allow patient's use of healthcare services and included income, state buy-in indicating whether a patient was enrolled in a state-administered Medicaid program, geographic location, and SEER region. Need characteristics determine the need for treatment or health services use and included presence of comorbid conditions, disability status, stage at diagnosis, tumor grade, and prior cancer history. The inclusion of need characteristics is particularly important in an environment where personalized medicine and risk-stratified approaches to disease management are increasingly adopted as ways to provide high-value, high-quality care. Patients' individual indication for treatment might also drive divergence of clinical

practice from established guidelines, thus challenging guideline concordance as a traditional quality-of-care indicator.

Data Sources

We used the National Cancer Institute's (NCI) Surveillance Epidemiology and End Result (SEER) database linked with Medicare fee-for-service claims from 2000 to 2014. The SEER data contain longitudinal demographic and incident cancer characteristics including grade and stage for approximately 28% of the U.S. cancer population. Data are available on all cancer cases diagnosed from 2000 and later for all SEER 18 registries participating in the program. The SEER 18 registries consist of the SEER 13 plus Greater California, Greater Georgia, Kentucky, Louisiana, and New Jersey ("SEER Registry Groupings for Analyses" 2017). Even though in 2001 the SEER Program expanded coverage to include Kentucky and the remaining counties in California and, in addition, New Jersey and Louisiana once again became participants, the first full year of data for all of these registries, part of SEER 18, is 2000 ("Greater California Registry" 2017; "Kentucky Registry" 2017; "Louisiana Registry" 2017; "New Jersey Registry" 2017). Hence, we had access to data from all 18 SEER registries for the study period.

Medicare provides health insurance for 97% of Americans aged 65 and older, and these data reflect health care services used as well as co-morbid health conditions (Warren et al. 2002). The Medicare inpatient file includes institutional claims submitted for facility costs covered under Medicare Part A as well as beneficiary, physician, and hospital identifiers, admission and discharge dates, and diagnosis and procedure codes. The outpatient files contain claims from outpatient institutional providers, such as hospital outpatient departments. The carrier files contain claims from non-institutional professional providers, such as physicians, nurse

practitioners, independent clinical laboratories, and ambulatory surgery centers. The pharmacy files include prescription drug data such as drug name, fill date, days supply, formulary status, and cost. Claims for Medicare beneficiaries enrolled in a Medicare managed care plan are not included (Warren et al. 2002).

To link SEER with Medicare data, the registries participating in the SEER program send individual identifiers for all persons in their files to NCI and the Centers for Medicare and Medicaid Services (CMS). These identifiers are matched with identifiers contained in Medicare's master enrollment file. The linkage was first completed in 1991 and has been updated in 1995, 1999, 2003, 2006, 2009, 2012, 2014, and 2016. For each of the linkages, 93 percent of persons aged 65 and older in the SEER files were matched to the Medicare enrollment file (Potosky et al. 1993; "SEER-Medicare: How the SEER & Medicare Data Are Linked" 2017). For this dissertation research we were granted access to the 2016 data linkage for the SEER 18 registries and thus we had claims data from 2000 to 2014 for Medicare beneficiaries with bladder cancer.

Study Population and Inclusion/Exclusion Criteria

The inclusion and exclusion criteria for the study cohorts for Aims 1 and 2 are presented in Table **5**. The study population included SEER-Medicare beneficiaries diagnosed with urothelial carcinoma of the bladder between January 1, 2001 and December 31, 2012. We used their Medicare claims from January 1, 2000 through December 31, 2014 to allow for identification of comorbidities and at least two years of follow-up claims post-diagnosis in order to assess outcomes. Patients aged 66 years and older were included in the study in order to have one year of claims prior to diagnosis to assess comorbidities and disability status (Klabunde, Warren, and Legler 2002; Klabunde et al. 2000; Davidoff et al. 2013).

We identified bladder cancer patients with urothelial non-muscle invasive (Ta, Tis, T1, N0, NX, M0) disease. Patients had to be continuously enrolled in Medicare parts A and B for a minimum of 12 months prior to first recorded diagnosis until the end of the study period or death. Patients were excluded if they were diagnosed at autopsy or death. We excluded patients who were enrolled in Medicare Advantage (managed care) from the year before diagnosis through the end of the study period or death because individual claims for those enrollees are not available (Warren et al. 2002). Patients were also excluded if they did not have a confirmed bladder cancer diagnosis (ICD-9) in any Medicare claims file within 2 months before or up until 12 months after the SEER diagnosis. Patients who underwent total cystectomy or died during the first two years post-diagnosis were censored, as they were no longer eligible for surveillance.

The study cohort for Aim 2 additionally excluded patients who died or underwent total cystectomy during the first two years after diagnosis as we could not assign them to a surveillance cystoscopy group, which requires two years of observed exposure.

Table 5. Study Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Aims 1 & 2	**Aims 1 & 2**
• All patients diagnosed with urothelial (histology codes 8050-8052, 8120-8124, 8130-8131), non-muscle-invasive (Ta, Tis, T1, N0, NX, M0) bladder cancer (SEER cancer codes C67.0-C67.9; ICD-9 codes 188.0-188.9, 233.7) from 2001-2012, allowing for at least two years of follow-up claims post-diagnosis in order to assess outcomes.	• Non-urothelial or missing histology • Diagnosed with advanced or metastatic cancer (T3-T4, N1-N3, M1) or missing stage • Diagnosed at autopsy or death
• Age 66 years or older at diagnosis in order to allow one year prior to diagnosis to assess comorbidities and disability status.	• <Age 66 at diagnosis
• Continuous Medicare Part A & B coverage for 12 months prior to diagnosis until end of the study period or death	• Lack of continuous Medicare Part A & B coverage until December, 2014 or death • Enrolled in Medicare Advantage (managed care)
• Bladder cancer diagnosis (ICD-9) in any Medicare claims file	• Lack of bladder cancer diagnosis on a Medicare claims file within 2 months before or 12 months after the SEER diagnosis
Aim 2 Only	**Aim 2 Only**
• Patients who survived for at least two years after diagnosis	• Patients who had died during the first two years post-diagnosis
• Patients who did not undergo total cystectomy during the first two years post-diagnosis	• Patients who had a total cystectomy during the first two years post-diagnosis

Diagnostic and procedural codes used to identify the cohort and relevant procedures and processes of care are presented in Appendices 3.1 and 3.2. Description of the American Joint Committee on Cancer (AJCC) TNM staging system for bladder cancer used to define the cohort is provided in Appendix 3.3. Corresponding anatomic stage/prognostic groups are described in Appendix 3.4.

Sample Size and Statistical Power

From 2000 onward there have been over 5,000 cases of bladder cancer every year available within the SEER-Medicare data (National Cancer Institute. Division of Cancer Control

& Population 2016). The study population for this dissertation, which includes patients diagnosed over twelve years from 2001 to 2012, would have an estimated 68,000 potential subjects. Of those, 70% are likely to be diagnosed with superficial (non-muscle-invasive) disease, for a total of approximately 47,000. Based on previous studies of patients with NMIBC, after applying standard inclusion/exclusion criteria about half of the original cohort is retained, therefore we estimate a final sample size of at least 23,500 patients who meet the study criteria (Schrag et al. 2003; Chamie et al. 2011; Chamie et al. 2015; Chamie et al. 2012; Hollingsworth et al. 2010). The expected sample size is similar to that found in prior studies examining treatment and surveillance use among NMIBC patients in SEER-Medicare data. The final cohort sizes from those studies ranged from 18,276 to 24,980 NMIBC patients diagnosed over ten years (Strope et al. 2010; Hollingsworth et al. 2010; Hollenbeck et al. 2009).

Using Cohen's criteria for assessing power, with an alpha level of 0.05 for a two-sided test of significance, and a minimum power of 80%, we should be able to detect small differences in means or proportions (Cohen 1992).

Variables and Measurement

Variable definitions, outcome windows, measures of interest, and sources are presented in Table **6** and Table **7**. The main outcome of interest in Aim 1 was receipt of surveillance cystoscopy. Main outcomes in Aim 2 were all-cause, bladder cancer-related, and other-cause mortality. Key explanatory variables for Aim 1 were patient-level characteristics, specifically: age at diagnosis, disability status and comorbidities measured in the year prior to diagnosis. The key explanatory variable for Aim 2 was receipt of cystoscopy.

Additional control variables based on a review of the literature and the conceptual model included: 1) predisposing characteristics: sex, race/ethnicity, marital status at diagnosis, census

zip code education level; 2) enabling characteristics: zip code median household income, Medicaid state buy-in indicator, geographic location, and SEER region; 3) need characteristics: stage at diagnosis, tumor grade, prior history of cancer; and 4) practice setting characteristics: surgeon volume.

Dependent variables

Receipt of cystoscopy

For aim 1, the primary outcome of interest was receipt of surveillance cystoscopy. Cystoscopy was identified using CPT codes on outpatient, carrier, and DME claims files using previously established CPT code sets, ICD-9 codes from inpatient files, and review of annual HCPCS coding manuals (Appendix 3.2). Receipt of recommended cystoscopy was defined based on the historic AUA guidelines (2000-2007) which were in effect during the study period (Hall et al. 2007). The new AUA/SUO guidelines recommending different intervals of surveillance based on patients' individual risks for recurrence and progression were only published in 2016 (Chang et al. 2016). Although guidelines endorsing a similar risk-stratified approach to surveillance were adopted at least eight years earlier by the European Association of Urology (EAU) (Babjuk et al. 2008), it would be difficult to ascertain whether physicians in the United States were influenced by the European recommendations and followed them instead of the domestic AUA guidelines. Table **1** presents a comparison of existing clinical guidelines for the diagnosis and management of NMIBC in the United States and Europe.

Receipt of recommended surveillance cystoscopy was defined as a binary variable indicating receipt of 7 or more cystoscopies over the first two years post diagnosis. From 2000 to 2007, the AUA guidelines for NMIBC recommended that for the first two years after diagnosis, at a minimum, patients should undergo lower urinary tract surveillance with cystoscopy every 3

months (Hall et al. 2007). Thus, if patients were completely compliant with these recommendations, they would have to undergo at least 8 cystoscopies over the first two years after diagnosis, a measure of compliance with surveillance used in previous studies (Chamie et al. 2011; Chamie et al. 2012). In practice, however, receipt of initial surveillance cystoscopy may be delayed, therefore we used a less stringent definition of receipt of recommended surveillance as 7 or more cystoscopies during the first two years after diagnosis. In sensitivity analysis we applied another definition of low-intensity surveillance as 4 or more cystoscopies over the first two years post-diagnosis to explore how sensitive the results are (Table **6**). We also used number of cystoscopies received as a dependent count variable in Aim 1 in order to model the outcome as count data and examine the marginal effect of the explanatory variables of interest on the predicted number of events (i.e., cystoscopies received).

Survival outcomes

For aim 2, the outcomes of interest were all-cause, bladder-cancer, and other-cause mortality. All-cause mortality was measured from January 1, 2001 through December 31, 2016 using the date of death from the Medicare Enrollment Database (EDB) (Bach et al. 2002). The Medicare date of death in the EDB is updated nightly from the Social Security Administration and is current as of the day that the enrollment data were extracted for the NCI's use (Bach et al. 2002). However, since their age and comorbidities places patients with NMIBC at risk for dying from competing causes, we also assessed bladder cancer-specific and other-cause mortality using the cause of death field from SEER. The SEER death date is primarily obtained from state death certificates (Bach et al. 2002) and was available from the SEER PEDSF through December, 2013. A recent analysis of six major cancer types evaluated in SEER showed that although the

utility of cause of death in calculating cancer-specific survival (CSS) depended on the risk of cancer-related mortality and non-tumor factors, the impact of this variation on CSS was small and the cause of death assigned by cancer registries has acceptable validity (Hu et al. 2013).

Figure **2** below illustrates the exposure period for receipt of cystoscopy and outcome time windows.

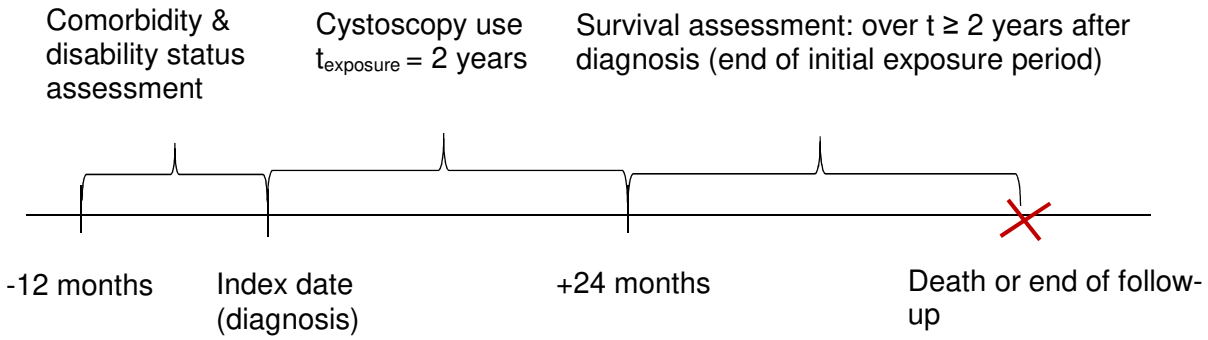


Figure 2. Exposure and outcome time windows

Table 6. Outcome definitions and time windows

Outcome	Definition	Anchor	Time Frame/ Follow-up	Specific Aim	Sensitivity analysis
Receipt of recommended cystoscopy based on the 2000-2007 AUA Guidelines	Binary variable for use of 7 or more cystoscopies over 2 years.	Diagnosis	2 years post-diagnosis	Aim 1	Binary variable for use of 4 or more cystoscopies over 2 years.
Number of cystoscopies	Count variable for number of cystoscopies received	Diagnosis	2 years post-diagnosis	Aim 1	N/A
Survival • All-cause mortality • Bladder cancer-specific mortality • Other-cause mortality		2 years post-diagnosis (end of exposure period)	Until death or end of follow-up	Aim 2	

Explanatory variables

The main explanatory variables included predisposing (demographic), enabling (i.e., income, state buy-in indicator, location, region), need (i.e., stage, tumor characteristics, disability status, comorbidities), and practice setting factors (surgeon volume) and were obtained from the SEER-Medicare linked data files (Table **7**). *Predisposing characteristics* included age at diagnosis, sex, race/ethnicity, marital status at diagnosis, and census zip code education level and were obtained from the SEER Patient Entitlement and Diagnosis Summary File (PEDSF) and the Zip Code Census File.

Race was defined as used in previous studies of surveillance and bladder cancer by Chamie and colleagues to allow comparability between studies (Chamie et al. 2011; Chamie et al. 2012). Race/ethnicity information was extracted from SEER PEDSF data instead of Medicare-reported race/ethnicity data in the EDB because of well-known measurement problems and inconsistencies over time in the SSA's definition of racial and ethnic groups. Furthermore, SEER uses Hispanic ancestry as an ethnic characteristic, captured by its own variable, in addition to a Hispanic-surname algorithm to ensure greater accuracy in reporting Hispanic ethnicity (Bach et al. 2002). For the purposes of this analysis, racial/ethnic groups were defined as white (non-Hispanic), black (non-Hispanic), and other (including Hispanic).

Age was modeled as a categorical variable using five groups (66-69; 70-74; 75-79; 80-84; ≥85) because the effect of age on receipt of cystoscopies likely differs across the age groups. A prior study found that being age 75 years or older was statistically significantly associated with receipt of low-intensity surveillance (Schrag et al. 2003).

The education variable was derived from the local census zip code characteristics and defined as proportion of residents in subject's zip code with less than a high school education, divided into quartiles.

Enabling characteristics included income, state buy-in indicator, geographic location, and SEER region. Income was also derived from the local census-tract characteristics and was defined as quartiles of median household income in subject's zip code. A history of any state buy-in in the 12 months preceding bladder cancer diagnosis was used as a proxy for individual low-income status. The state buy-in variable generally indicates whether a beneficiary was enrolled in a state-administered Medicaid program. A validation study showed, however, that the state buy-in variable alone has low sensitivity and may be inadequate to identify dual Medicare-Medicaid beneficiaries (Koroukian et al. 2010). Geographic location was defined as a categorical variable indicating extent of urbanization of patient's residence at the time of diagnosis (big metro, metro, urban, less, urban, rural). Descriptions of each category are provided in Table **7**. Region was defined as a categorical variable grouping the SEER registries in four main regions (Northeast, West, Midwest, South).

Need characteristics included stage at diagnosis, tumor grade, comorbidities and disability status in the year prior to diagnosis. Patient comorbidities were assed using the Klabunde modification of the Charlson comorbidity index based on patients' Medicare Parts A and B claims pre-diagnosis (Klabunde et al. 2000). Patients were classified as having 0, 1, 2, or 3 and more comorbidities.

Disability status (DS) was used as an additional key explanatory variable to reduce confounding by indication. A main limitation of previous studies on surveillance among bladder cancer patients is the lack of critically important information about patient performance and functional status, both of which may influence patient and provider thresholds for performing regular bladder surveillance (Schrag et al. 2003; Chamie et al. 2011). We assessed DS using a novel claims-based prediction algorithm for performance status (PS) in older adults developed by

Davidoff and colleagues (Davidoff et al. 2013). DS was defined as a dichotomous indicator of good/poor predicted DS at diagnosis (Davidoff 2014).

The DS measure was developed by Davidoff and colleagues using data from the 2001–2005 Medicare Current Beneficiary Survey (MCBS) and a representative sample of Medicare beneficiary population ages 66 and over. Responses to the various self-reported measures of functional status, strength, stamina, and exercise on the MCBS were used to construct a proxy measure for the 6-level Eastern Cooperative Oncology Group (ECOG) Performance Status (PS) scale and the initial measure was subsequently collapsed into a dichotomous indicator for good (0-2) versus poor (3-4) DS (Davidoff et al. 2013). 5 on the ECOG PS scale indicates death. Davidoff et al. developed a claims-based prediction algorithm including various indicators for health care services in the 12 months prior cancer diagnosis that were expected to vary based on DS level (Davidoff et al. 2013). The model captured relationships between healthcare utilization and functional status, independent of the effects of cancer and its treatment, which is consistent with the approach used in the ECOG PS scale, and appropriate to capture baseline health status at the point of cancer diagnosis (Davidoff et al. 2013). The DS measure was implemented and validated in 4 cohorts of cancer patients: early and advanced non-small cell lung cancers (NSCLC), stage IV estrogen-receptor negative breast cancer, and myelodysplastic syndromes (MDS) and was shown to be a significant independent predictor of cancer-directed treatment among the four cancer cohorts but has not been applied to bladder cancer population before (Davidoff 2014).

We used the model with permission from Dr. Davidoff for the purposes of this dissertation and specific research questions (email correspondence with Dr. Davidoff from 2/21/2017).

Practice setting characteristics: since surveillance cystoscopy is typically administered in an outpatient setting and does not require hospitalization, we were not able to examine the association between hospital characteristics and patient surveillance. However, the structural elements that we examined were distance traveled to cystoscopy provider and surgeon case volume. Surgeon volume was defined as the number of transurethral resections performed by each patients' primary treating provider in the year prior to patient's diagnosis. Surgeons with higher volume of transurethral resections might be more likely to use more frequent surveillance cystoscopy due to inherent preferences or proclivity for patient workup and endoscopic evaluation. Therefore, it is important to account for surgeon volume in the multivariable models examining factors associated with receipt of surveillance cystoscopy. Distance traveled to cystoscopy provider was calculated from subject's zip code to the zip code of the primary cystoscopy identified in the Medicare claims data.

Table 7. Variable measures and data sources

Level	Variable	Definition	Variable Type	Referent Category	Specific Aim	Data source: SEER-Medicare
Key Dependent Variables						
Health services utilization measures	Receipt of surveillance cystoscopy	1) ≥7 cystoscopies over 2 years 2) ≥4 cystoscopies over 2 years 3) Number of cystoscopies over 2 years	Binary Count	1) <7 over 2 years 2) < 4 over 2 years 3) 0	Aim 1	Outpatient, carrier, and DME claims
Long-term health outcomes	Survival	1) All-cause mortality 2) Bladder cancer-specific mortality 3) Other-cause mortality	Continuous		Aims 2 & 3	PEDSF
Economic outcomes	Costs (Medicare payment and patient out-of-pocket costs)	1) Cystoscopy 2) Cystectomy 3) TURBT	Continuous		Aim 3	outpatient, carrier, inpatient, DME claims
Key Explanatory Variables ***(SEER-Medicare variable name italicized in parentheses)***						
Predisposing characteristics	Age at diagnosis *(agedx)*	66-69; 70-74; 75-79; 80-84; ≥85	Categorical	66-69	Aims 1-3	PEDSF
	Sex *(m_sex)*	1=female; 0=male	Binary	Male	Aims 1-3	PEDSF
	Race/ethnicity *(race)*	White (non-Hispanic); Black (non-Hispanic); Other (including Hispanic)	Categorical	White (non-Hispanic)	Aims 1 & 2	PEDSF
	Marital status at diagnosis *(marst)*	Married/domestic partner; single; other (divorced, widowed, separated); unknown	Categorical	Married/ domestic partner	Aims 1 & 2	PEDSF
	Census zip-code education level *(ctnon)*	Proportion of residents in subject's zip code with less than a high school education	Categorical, quartiles: 1 (low) – 4 (high)	1 (low)	Aims 1 & 2	Zip Code Census File

Enabling characteristics	Census zip-code income level *(ctmed)*	Quartiles of median household income in subject's zip-code	Categorical, quartiles: 1 (low) – 4 (high)	1 (low)	Aims 1 & 2	Zip Code Census File
	State buy-in indicator *(allflag)*	History of state buy-in during the year preceding diagnosis	Binary 1=any state buy-in; 0=no buy-in	No buy-in	Aim 1	PEDSF
	Geographic location *(urbrur)*	Urban vs rural residence	Categorical: 1 = Big Metro (≥1,000,000 metropolitan pop) 2 = Metro (< 1,000,000 metropolitan pop) 3 = Urban (≥20,000 nonmetropolitan pop) 4 = Less Urban (2,500-19,999 nonmetropolitan pop) 5 = Rural (Completely rural or less than 2,500 urban pop) 9 = Unknown	Urban	Aims 1 & 2	PEDSF
	SEER registry region *(reg)*	West, Midwest, South, Northeast	Categorical	South	Aims 1 & 2	PEDSF
Need characteristics	Stage at diagnosis (*t_value – AJCC 3rd ed, 2001-2003; dajcct – AJCC 6th ed, 2004+)*	Ta, Tis, T1	Categorical	Ta	Aims 1-3	PEDSF
	Tumor grade *(grade)*	G1 - Well differentiated; G2 - moderately differentiated; G3 - poorly differentiated; G4 – undifferentiated; Unknown	Categorical	Well differentiated	Aims 1-3	PEDSF
	Comorbidities at diagnosis	Charlson comorbidity index (0, 1, 2, or ≥3	Categorical	0	Aims 1 & 2	outpatient, carrier, inpatient, DME,

		comorbid conditions)				hospice, home health claims
	Disability status (DS) at diagnosis	Predicted DS 1=poor DS 0=good DS	Binary	Good DS	Aims 1 & 2	outpatient, carrier, inpatient, DME, hospice, home health claims
	Prior history of cancer	1=prior history 0=no history	Binary	No history	Aims 1 & 2	PEDSF
Practice setting characteristics	Distance traveled	Travel distance to cystoscopy provider calculated from unencrypted patient and provider zip codes	Categorical (quartiles)	1 (low)	Aims 1	PEDSF, carrier (NCH physician/ supplier) file
	Surgeon volume	Number transurethral resections performed by the primary provider in the 12 months prior to diagnosis	Count	0	Aim 1	Outpatient, DME, and carrier files
Additional controls	Time *(yrdx)*	Year of diagnosis	Categorical	2001	Aims 1 & 2	PEDSF

Statistical Analyses by Aim

Aims 1 and 2 were retrospective cohort studies of Medicare beneficiaries diagnosed with bladder cancer from 2001 to 2012. Aim 3 used a patient-level simulation model calibrated with data from SEER-Medicare, clinical trials, and the published literature to compare the cost-effectiveness of the US risk-stratified guidelines with the historic uniform recommendations. Research questions, hypotheses, and analyses are summarized by aim in Table **8**.

Analyses were performed using SAS® Studio software, Version 3.7 (Enterprise Edition) for Linux (SAS Institute, Inc., Cary, NC, USA), Stata Statistical Software: Release 14

(StataCorp. 2015, StataCorp LP, College Station, TX, USA), and TreeAge Pro 2017, R2.1 (TreeAge Software, Inc., Williamstown, MA, USA).

Table 8. Proposed research questions, hypotheses, and analyses summarized by aim

Hypotheses		**Model**	**Analysis**
AIM 1 – Q1: What are the surveillance patterns and which NMIBC patients are more likely to receive recommended surveillance cystoscopy within Medicare?			
H1a	Older patients with multiple comorbidities and poor disability status will be *less likely* to have received recommended surveillance cystoscopy (≥7 v <7; ≥4 v <4), ceteris paribus.	Multilevel (mixed effects) logistic regression	$Pr(Cysto) = f(\beta_0 + \beta_1 Age + \beta_2 Comorbidity + \beta_3 Disability + \beta_4 Controls + \beta_6 Year)$
H1b	Older patients with multiple comorbidities and poor disability status will be *more likely* to have received *fewer number* of cystoscopies, ceteris paribus.	Multilevel (mixed effects) Poisson regression	$E(Cysto \mid X_{ij}) = exp(\beta_0 + \beta_1 Age + \beta_2 Comorbidity + \beta_3 Disability + \beta_4 Controls + \beta_6 Year)$
AIM 2 – Q2: Is receipt of surveillance cystoscopy associated with improved survival within the Medicare NMIBC population?			
H2a	Surveillance cystoscopy use will not be associated with significant improvement in overall survival, after controlling for patient demographic and clinical characteristics.	Propensity score-weighted Cox proportional hazard model	$h(z \mid X_{ij}) = h_0(z)\, exp(\beta_1 Cysto)$
H2a	Surveillance cystoscopy use will not be associated with significant improvement in bladder cancer-specific survival, after controlling for patient demographic and clinical characteristics.	Propensity score-weighted competing-risks regression model	$h(z \mid X_{ij}) = h_0(z)\, exp(\beta_1 Cysto + \beta_2 CompetingRisks)$
AIM 3 – Q3: What is the cost-effectiveness of three different surveillance cystoscopy strategies over the first 5 years post-diagnosis among older NMIBC patients?			
H3a	Low-intensity risk-stratified surveillance will incur *fewer costs per QALY gained* compared with other surveillance approaches.	Patient-level simulation model	ICER planes, CEACs
H3b	Low-intensity risk-stratified surveillance will incur *fewer costs per recurrence detected* compared with other surveillance approaches.	Patient-level simulation model	ICER planes, CEACs
H3c	Low-intensity risk-stratified surveillance will incur *fewer costs per progression averted* compared with other surveillance approaches.	Patient-level simulation model	ICER planes, CEACs
H3d	Low-intensity risk-stratified surveillance will incur *fewer costs per death averted* compared with other surveillance approaches.	Patient-level simulation model	ICER planes, CEACs

Aim 1: Examine surveillance patterns and factors associated with receipt of recommended surveillance cystoscopy for NMIBC patients.

Hypothesis 1a: Patient comorbidities, age, and disability status will be negatively associated with receipt of recommended surveillance cystoscopy (≥ 7 cystoscopies over two years post diagnosis), after controlling for demographic and clinical characteristics.

Hypothesis 1b: Patient comorbidities, age, and disability will be negatively associated with the *number* of cystoscopies received, after controlling for demographic and clinical characteristics.

Approach:

Descriptive statistics

Descriptive statistics were calculated for each variable as percentages for categorical variables and means/standard deviations for continuous variables. We compared baseline characteristics between patients with poor and good DS using Chi-square test for categorical variables and two-sided Student's t test for continuous variables.

Multivariable models

1) *Two-level mixed effects logistic regression*

To test hypothesis 1a, we used two-level logistic regression models for our binary outcomes (≥7 or ≥4 cystoscopies). All multivariable models accounted for clustering of patients within physicians and allowed the intercept term to vary randomly over physician groups (T. A. B. Snijders and Bosker 2012; Curran and Bauer 2007) The models also included an offset term to account for the number of days each patient was observed and "at risk" for surveillance until

death or total cystectomy. Patients who underwent total cystectomy were censored, as they were no longer eligible for surveillance as they had their bladder removed.

Characteristics of the individual patients represented the level 1 ("fixed") categories of the model and the provider who performed the cystoscopy (represented by the UPIN or NPI) represented the level 2. Unique Physician Identifier Numbers (UPINs) are six-character identifiers used by CMS to identify physicians accepting Medicare insurance in the United States. UPINs were discontinued in June 2007 and replaced by National Provider Identifier (NPIs) (CMS 2012).

A previous study examining management of NMIBC patients in SEER-Medicare between 1992 and 2002 identified 4,545 patients nested within 1,536 provider practices, therefore we expected to have enough patients clustered within physicians to render the analysis feasible (Chamie et al. 2011).

In multilevel models the variables for which there is no wider population of categories to sample from (e.g., age, ethnicity, sex) are considered fixed categories. Variable for which the units can be regarded as a random sample from a wider population of units (e.g., physicians, hospitals) are considered the levels. The higher the level, the fewer the number of units at that level (e.g., physicians within hospitals) (T. A. Snijders and Bosker 2012). An advantage of a multilevel model compared to other multivariable regression techniques is that it accounts for the hierarchical structure of the data (patients nested within physicians nested within hospitals) and therefore produces correct standard errors (SEs) which do not need to be adjusted ex-post using clustered SEs. In maximum likelihood estimations of logit models an ex-post adjustment to SEs can produce not only biased SEs but also biased coefficient estimates.

Assigning patients to treating physicians:

We identified the physician who performed the initial cystoscopy for each patient using the UPIN or NPI included in the Medicare claims files. When a patient had more than two procedures, we assigned that patient to the provider who performed the greatest number of procedures. Following prior studies' method for assigning bladder cancer patients to providers (Schrag et al. 2003), if a patient had two urologic procedures performed by two different providers, we selected the provider who performed the second procedure, under the assumption that the provider who had the most recent contact with the patient is likely to exert the greatest influence on that patient's surveillance behavior.

We estimated the following mixed-effects model of receipt of recommended cystoscopy:

$$\Pr(CYSTO_{ijt}) = f(\gamma_{00} + X_{ij}\gamma_{10} + Z_{ij}\gamma_{20} + W_{ij}\gamma_{30} + V_j\gamma_{01} + X_{ij}u_{1j} + Z_{ij}u_{2j} + W_{ij}u_{3j} + \delta_t + u_{0jt} + \varepsilon_{ijt}) \qquad (1)$$

where $CYSTO_{ijt}$ is the dependent variable indicating receipt of recommended cystoscopy, X_{ij} is a vector capturing predisposing characteristics (patient demographics), Z_{ij} is a vector capturing enabling characteristics (e.g., income; state buy-in; location), W_{ij} is a vector capturing need characteristics (clinical factors, DS, comorbidities), and V_j is a vector capturing practice setting characteristics (surgeon volume). δ_t is year of diagnosis fixed effects to account for any time trends in use of cystoscopy. The segment $[\gamma_{00} + X_{ij}\gamma_{10} + Z_{ij}\gamma_{20} + W_{ij}\gamma_{30} + V_j\gamma_{01}]$ in equation 1 contains the fixed coefficients and represents level 1 of the model. The segment $[X_{ij}u_{1j} + Z_{ij}u_{2j} + W_{ij}u_{3j} + \delta_t + u_{0jt} + \varepsilon_{ijt}]$ contains the random error terms and accounts for clustering at the physician level.

Alternative approaches / sensitivity analyses

Mixed-effects (multilevel) logistic regression models have been commonly used in previous studies examining factors associated with receipt of treatment or surveillance among bladder cancer patients (Chamie et al. 2011; Chamie et al. 2012; Hollingsworth et al. 2010; Gore et al. 2010). An advantage of the mixed-effects logistic model over a RE logit model is that the mixed effects model can generate predictions or marginal effects for all those in the sample, while the RE logit model can generate predictions for those individuals whose random effect equals a specific value, which is not representative of the full sample. Furthermore, a key assumption of the RE models is that the individual specific effect is not correlated with explanatory variables in the model (Cameron and Trivedi 2010).

An alternative approach in the presence of convergence problems with the mixed-effects logistic regression model or an insufficient number of physician clusters, is using a generalized estimating equations (GEE) model with a binomial distribution (family), logit link, and exchangeable correlation. The main difference between GEE and multilevel or RE logit models is that unlike multilevel or RE models, which are "subject-specific" and allow to investigate changes in individuals' responses, GEE are a "population-averaged" approach and model marginal distributions. GEE models require to specify family, link, and a correlation structure across observations within each group. If the correct correlations structure is specified, GEE account better for clustering than a logit model with clustered standard errors because they take into account the specified correlation structure (e.g., exchangeable, unstructured) when selecting the final parameter estimates (Cameron and Trivedi 2010). However, GEE are robust to misspecification of the working correlation only when the number of clusters is sufficiently large (Panageas et al. 2003).

2) Two-level mixed-effects Poisson count data model

To test hypothesis 1b, we use a two-level Poisson model:

$$E(CYSTO_{ijt}|\, \alpha_{ij}, X_{ij,} Z_{ij,} W_{ij,} V_{ij,}) = \alpha_{ij} \exp(X_{ij}\gamma_{10} + Z_{ij}\gamma_{20} + W_{ij}\gamma_{30} + V_j\gamma_{01} + X_{ij}u_{1j} + Z_{ij}u_{2j} + W_{ij}u_{3j} + \delta_{t)} \quad (2)$$

where $CYSTO_{ijt}$ is the dependent variable indicating number cystoscopies received, X_{ij} is a vector capturing predisposing characteristics (patient demographics), Z_{ij} is a vector capturing enabling characteristics (e.g., income; state buy-in; location), W_{ij} is a vector capturing need characteristics (clinical factors, DS, comorbidities), and V_j is a vector capturing practice setting characteristics (e.g., distance travelled, surgeon volume). δ_t is year of diagnosis fixed effects to account for any time trends in use of cystoscopy. The models also included an offset term to account for the number of days each patient was observed and "at risk" for surveillance until death or total cystectomy. Patients who underwent total cystectomy were censored, as they were no longer eligible for surveillance as they had their bladder removed. A key assumption of the Poisson model is that of equidispersion (the variance equals the mean).

Alternative approaches / sensitivity analyses

While mixed-effects Poisson estimators account for clustering at the physician level in addition to the panel nature of the data, they are computationally intensive and may have convergence problems (Cameron and Trivedi 2010). Should the mixed-effects Poisson count data model have failed to converge, we would have used the following alternative approaches:

1) A random effects (RE) Poisson model with clustered standard errors.

a. The standard Poisson RE estimator assumes that α_i is gamma distributed with a mean of 1 and variance of η but the η parameter is not flexible enough to account for both overdispersion and correlation, therefore cluster-robust standard errors should be used to correct for that (Cameron and Trivedi 2010).

2) GEE with a Poisson distribution (family), log link, and exchangeable correlation.

Testing the fit of the Poisson models

A key assumption of the Poisson model is that of equidispersion (the variance equals the mean). However, count data are often overdispersed (the variance is larger than the mean).

To test the fit of the model, we used the Pearson goodness-of-fit test which compares the fitted values from the model to actual values. If the null hypothesis of a Poisson distribution was rejected, we would use a negative mixed effects or RE negative binomial model which relaxes the assumption that the variance equals the mean. The negative binomial estimator is designed to explicitly address both over-dispersion and within correlation, thus leading to more efficient estimates than those obtained using Poisson RE with cluster-robust SEs (Cameron and Trivedi 2010). However, an advantage of the Poisson panel estimators is that they rely on weaker distributional assumptions than the negative binomial estimators and may be more robust as a result (Cameron and Trivedi 2010).

Aim 2: Examine the association between surveillance cystoscopy use and NMIBC survival within the Medicare population.

Hypothesis: Surveillance cystoscopy use will not be associated with significant improvement in survival, after controlling for patient demographic and clinical characteristics.

Approach:

Propensity Score Estimation and Application

We estimated propensity scores by modeling the probability of receiving surveillance (≥7 cystoscopies or ≥4 cystoscopies) in the first two years following bladder cancer diagnosis as a function of all explanatory variables using two-level logistic regression models, accounting for clustering of patients within physicians. The second level of the two-level model was represented by the provider identification number and allowed the intercept term to vary randomly over physician groups (T. A. B. Snijders and Bosker 2012; Curran and Bauer 2007). The models adjusted for patients' age at diagnosis, sex, race/ethnicity, marital status, comorbidities and disability status at diagnosis, prior history of cancer, stage at diagnosis, tumor grade, history of Medicaid state buy-in in the year prior diagnosis, SEER region, zip-code level education and median household income, residential status at diagnosis, surgeon volume, and year of cancer diagnosis.

We used a post-estimation function from the two-level logistic regression models to generate propensity scores and created two sets of inverse probability of treatment weights (IPTW) for each patient: a) equal to the inverse of the propensity score [1/p] for patients who received ≥7 cystoscopies and ≥4 cystoscopies; and b) equal to the inverse of one minus the propensity score [1/(1 − p)] for the two control groups (patients who received <7 and <4 cystoscopies), based on the two different surveillance definitions.

We stabilized the propensity score weights by multiplying the treatment and comparison IPTW by a constant, equal to the expected value of being in the treatment or comparison groups (Robins, Hernán, and Brumback 2000; Garrido et al. 2014; Harder, Stuart, and Anthony 2011). The purpose of using stabilized weights is to reduce the weights of either those treated subjects

with low propensity scores or those untreated subjects with high propensity scores (S. Xu et al. 2017). This method of propensity score weighting provides an estimate of the treatment effect in the population (in this case, the effect of surveillance cystoscopy use among all NMIBC patients) (Stürmer, Rothman, and Glynn 2006).

For before and after matching comparisons between the treatment and control groups, we examined the standardized difference in the means (SDM) which is preferred to t-tests because SDM is not influenced by sample size (P. Austin 2009; P. C. Austin 2011). If the absolute difference is <10%, the two groups are considered balanced (Normand et al. 2001).

Survival Analysis

To characterize the association between bladder cancer-specific mortality and surveillance we used a propensity-score adjusted, Fine-Gray competing-risks regression model (Fine and Gray 1999). Failure was defined as bladder cancer mortality and the competing event as non-bladder cancer mortality. Estimates from the competing-risk models were reported as subdistribution hazard ratios (SHR) with corresponding 95% confidence intervals (95% CIs) (Fine and Gray 1999). The explanatory variable included in the propensity-score weighted survival models was an indicator for receipt of cystoscopy (≥7 vs. <7 or ≥4 vs. <4 cystoscopies over the first two years post-diagnosis).

We used a propensity-score weighted Cox proportional hazards model to examine the association between all-cause mortality and surveillance where failure was defined as death from any cause. All models included standard errors clustered on the provider using the Huber-White sandwich variance estimator, because patients treated by the same provider may have similar outcomes. We confirmed the non-violation of the proportional hazards assumption (i.e., constant

hazard ratio over time) using Schoenfeld's residuals test and the time-dependent covariate test for interactions between time and the cystoscopy indicators (Bradburn et al. 2003a; Bradburn et al. 2003b).

As a sensitivity analysis we estimated bladder cancer-specific and other-cause mortality using Cox proportional hazards model for cause-specific mortality where failure was defined as death from bladder cancer, while the competing risks (other-cause deaths) were treated as censored events. A major advantage of the Fine-Gray model over the Cox proportional hazards model, especially if competing risks are frequent, is that the Fine-Gray model allows for dependency between the modeled competing risks (P. C. Austin, Lee, and Fine 2016). It has been also suggested in the literature that the cause-specific Cox proportional hazard model is preferred when studying disease etiology while the Fine-Gray subdistribution hazard model is preferred when estimating individual's risk and prognosis or the overall impact of covariates on the incidence of the outcome of interest (P. C. Austin, Lee, and Fine 2016).

In sensitivity analyses, we compared results from the propensity-score weighted survival models using stabilized inverse probability of treatment weights (SIPTW) to those from multivariable adjusted models (including separately all covariates from the propensity score model in addition to the cystoscopy indicators) and unadjusted models (including cystoscopy indicators only).

Lastly, we generated cumulative incidence functions (CIFs) from the propensity-score weighted Fine-Gray competing risk regression to compare the cumulative incidence of bladder-cancer or other-cause death between patients who received ≥7 vs. <7 and ≥4 vs. <4 cystoscopies. We also stratified the CIFs by key covariates, including age at diagnosis, disability status, and comorbidities.

Limitations / alternative approaches

A propensity score (PS) is defined as the probability of exposure (e.g., treatment), given a set of observed (measurable) characteristics in the data (Rosenbaum and Rubin 1983). By matching exposed and unexposed subjects based on observed covariates, PS methods can help reduce the bias in estimating treatment effects and the likelihood of confounding when analyzing nonrandomized, observational data (Haukoos and Lewis 2015). A major limitation of propensity score methods is that unadjusted confounding may still exist if unmeasured or unobserved factors, not included in the PS model, influenced treatment selection. A review of the application of PS methods in 69 studies found that whether PS methods or conventional multivariable models were used to control for confounding had little effect on results in those studies in which such comparison was possible. The authors concluded that there is little evidence that PS methods yield substantially different estimates compared with conventional multivariable methods (Stürmer et al. 2006).

A better way to account for potential unmeasured sources of confounding and selection bias would be to use an instrumental variable (IV) analysis. A prerequisite for IV analysis is to identify a strong IV which can introduce exogenous variation and is strongly correlated with the outcome of interest (X) but is not directly correlated with the treatment Y given X. A major challenge to identifying IVs is meeting both the requirements of strong correlation with X and the exclusion restriction. Distance traveled to cystectomy provider has been used in previous observational studies of survival in patients with bladder cancer and head and neck cancers and was found to be related to treatment but not directly related to survival outcomes (Gore et al. 2010; Beadle et al. 2014). A similar approach would be to use distance to cystoscopy provider as an IV. However, unlike cystectomy, cystoscopy is a common and easily accessible procedure.

Therefore, distance to cystoscopy provider is unlikely to be strongly correlated with receipt of cystoscopy, thus yielding a weak instrument.

***Note on detecting recurrence and/or progression in claims data*:**

The SEER registries do not conduct active follow-up of patients or collect information on recurrence, progression, or metastasis occurring after initial diagnosis. As a result, there are challenges estimating accurately progression-free survival in SEER-Medicare. Using procedural codes only to identify patients who received treatment for their progression will misclassify a large proportion of patients. Studies have shown that algorithms based on procedural codes to identify patients in SEER-Medicare with relapse or later metastatic disease have low sensitivity and are likely to miss a large percentage of patients with recurrences, particularly those who are older, and result in an incompletely and inaccurately classified cohort (Nordstrom et al. 2016; Earle et al. 2002; Warren et al. 2016; Warren and Yabroff 2015). For those reasons, we did not assess progression-free or recurrence-free survival in SEER-Medicare.

Aim 3: Assess the cost-effectiveness of the US risk-stratified surveillance guidelines compared with the historic recommendations of uniform surveillance among Medicare patients with NMIBC.

Hypothesis: Risk-stratified surveillance (intensive surveillance only among patients at high risk of recurrence and less frequent surveillance among those at low or intermediate risk) will be associated with similar survival but lower costs compared with historic recommendations

of cystoscopy every three months for two years and less frequently thereafter for all NMIBC patients, regardless of risk.

Approach:

Overview

We developed a patient-level simulation model using TreeAge Pro 2017, R2.1 (TreeAge Software, Inc., Williamstown, MA, USA) to compare three different surveillance strategies: historic AUA guidelines, low-intensity AUA/SUO risk stratified approach, and high-intensity AUA/SUO risk-stratified approach (Table **9**) (Hall et al. 2007; Chang et al. 2016). The model had a time horizon of 5 years with three-month cycles. We projected downstream outcomes and costs for a hypothetical closed cohort of 100,000 Medicare patients with non-muscle invasive bladder cancer (NMIBC). Patients were stratified in three risk groups for recurrence and progression based on the 2016 AUA/SUO guidelines (Table **2**).

The structure and flow of the model are outlined in Figure **3**. Patients aged ≥66 years entered the "disease free" state following an initial transurethral resection of the bladder tumor (TURBT). At each 3-monthly model cycle the patient could experience a bladder cancer recurrence. If the recurrence was detected, the patient underwent an additional TURBT and returned to a disease free state. However, if the recurrence was not detected, then the patient was at risk of progression and would undergo further treatment (cystectomy) once this progression was eventually detected. Death from background mortality could occur from any health state and bladder cancer mortality could occur only from the progression state or post-cystectomy state.

Costs and outcomes (quality-adjusted life years) were discounted at an annual rate of 3%. Costs were evaluated from the health care sector and societal perspectives (Sanders et al. 2016). All costs were inflated to 2017 USD, using the Consumer Price Index (CPI) for medical care.

Model Assumptions

We made the following assumptions, following prior decision modeling work on bladder cancer surveillance and review of the literature (Appendix 2.2):

1) Only one recurrence could develop per surveillance cycle.
2) Progression could occur only after recurrence (De Bekker-Grob et al. 2009; Van Kessel et al. 2013).
3) Low-risk patients who experienced initial recurrence were re-classified as intermediate-risk (Chang et al. 2016). Patients who started as intermediate- or high-risk remained in the same risk group for the duration of the five-year model time horizon.
4) If the tumor progressed to muscle invasive bladder cancer and was detected, patients received cystectomy (De Bekker-Grob et al. 2009; Van Kessel et al. 2013).
5) After cystectomy, patients were no longer eligible for or underwent further surveillance (De Bekker-Grob et al. 2009; Van Kessel et al. 2013).
6) Only patients with tumor progression at a certain moment or those in the post-cystectomy state were at higher risk for death than the background mortality rate (De Bekker-Grob et al. 2009; Van Kessel et al. 2013).
7) Surveillance stopped after 5 years, since per the 2016 AUA/SUO guideline recommendations, the decision whether to continue or stop routine follow-up cystoscopy

after five years should be based on shared-decision making between the patient and clinician (Chang et al. 2016).

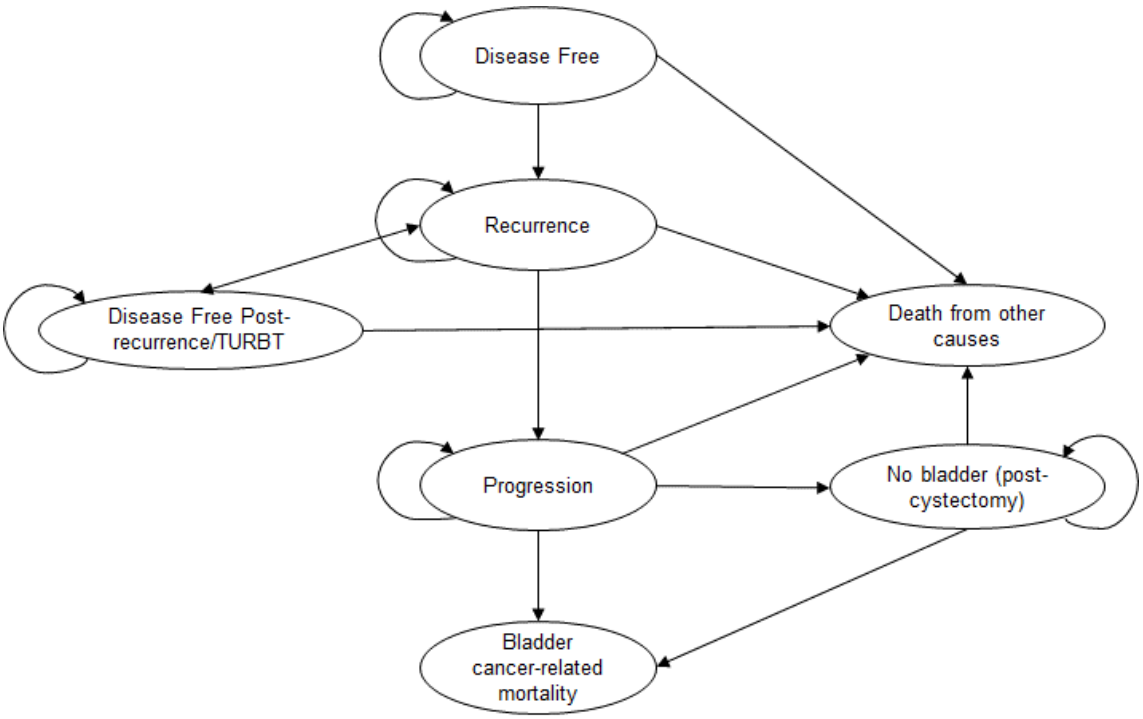


Figure 3. Disease state transition diagram.

Patients enter the "disease free" state following an initial transurethral resection of the bladder tumor (TURBT). At each 3-monthly model cycle the patient may experience a bladder cancer recurrence. If the recurrence is detected, the patient will undergo a further TURBT and return to a disease free state. However, if the recurrence is not detected, then the patient will be at risk of progression and will have to undergo further treatment (cystectomy) once this progression is eventually detected. Death from background mortality can occur from any health state and bladder cancer mortality can occur only from the progression state or post-cystectomy state.

Surveillance Strategies

We compared surveillance under the 2016 AUA/SUO risk-stratified guidelines (Chang et al. 2016) and the historic AUA guidelines (Hall et al. 2007). According to the 2016 AUA/SUO guidelines, for a low-risk patient whose first surveillance cystoscopy is negative for tumor, a clinician should perform subsequent surveillance cystoscopy six to nine months later, and then annually thereafter; surveillance after five years in the absence of recurrence should be based on

shared-decision making between the patient and clinician. For an intermediate-risk patient whose first surveillance cystoscopy is negative for tumor, a clinician should perform subsequent cystoscopy with cytology every 3-6 months for 2 years, then 6-12 months for years 3 and 4, and then annually thereafter. For a high-risk patient whose first surveillance cystoscopy is negative for tumor, a clinician should perform subsequent cystoscopy with cytology every three to four months for two years, then six months for years three and four, and then annually thereafter (Table **1**). The AUA/SUO definitions for risk-stratification in three groups (low, intermediate, and high) based on patients' risk for recurrence and progression are described in Table **2**.

We modelled three surveillance strategies based on the above definitions: low- and high-intensity risk-stratified recommendations and historic uniform guidelines (no risk-stratification) Table 9). Specifically, we modelled:

1) Low-intensity risk-stratified approach:
    a. Low-risk group: surveillance cystoscopy at three months, nine months, and then annually thereafter for five years.
    b. Intermediate-risk group: surveillance cystoscopy at three months, every six months for two years, and then annually thereafter;
    c. High-risk group: surveillance cystoscopy every three months for two years, every six months for years 3 and 4, and then annually thereafter;
2) High-intensity risk-stratified approach:
    a. Low-risk group: surveillance cystoscopy at three months, six months, and then annually thereafter for five years;
    b. Intermediate-risk group: surveillance cystoscopy every three months for two years, every six months for years 3 and 4, and then annually thereafter;

c. High-risk group: surveillance cystoscopy every three months for two years, every six months for years 3 and 4, and then annually thereafter;

3) Historic uniform guidelines: cystoscopy every 3 months for 2 years, every 6 months for 2 years, and annually thereafter.

Table 9. Surveillance strategies compared in the cost-effectiveness model

	Historical AUA Guidelines (2000-2007)	**Low-intensity AUA/SUO Risk-stratified Approach (2016)**			**High-intensity AUA/SUO Risk-stratified Approach (2016)**		
	All	Low Risk	Intermediate Risk	High Risk	Low Risk	Intermediate Risk	High Risk
Cystoscopy	Every 3 months for 2 years	3 months	3 months	Every 3 months for 2 years	3 months	Every 3 months for 2 years	Every 3 months for 2 years
	Every 6 months for 2 years	9 months	Every 6 months for 2 years	Every 6 months for years 3 and 4	6 months	Every 6 months for years 3 and 4	Every 6 months for years 3 and 4
	once/year	once/year	once/year	once/year	once/year	once/year	once/year
Total follow-up	5 years	5 years	5 years	5 years	5 years	5 years	5 years
Total cystoscopies	13	6	7	13	6	13	13

Model Outcomes

Model outcomes per patient and for the entire cohort were assessed over five years of surveillance under each strategy and included total quality-adjusted life years (QALYs), total cancer recurrences detected, total progressed cases detected, total deaths, total false positive (FP) cases, and total discounted costs (from the health care sector and societal perspectives).

Transition Probabilities, Cost, and Utility Data

Model parameters, values, and proposed distributions are presented in Table **13**. Data on patient characteristics used to stratify patients into the three risk groups (Table **2**) such as stage at

diagnosis, and tumor grade were obtained from the SEER-Medicare PEDSF files (Aims 1 & 2). Data on tumor size and presence of solitary vs. multifocal tumors were obtained from the EORTC trials (Sylvester et al. 2006). Remaining characteristics used in the 2016 AUA/SUO risk-stratification algorithm (e.g., any variant histology, any lymphovascular invasion; any high grade prostatic urethral involvement) were informed by review of the published literature.

Three-month state transition probabilities were obtained using conversion of the annual probabilities of death, recurrence, and progression (Fleurence and Hollenbeak 2007). Annual probabilities of recurrence and progression by year for five years for each risk group were obtained from the EORTC risk tables (Sylvester et al. 2006). Bladder cancer and other cause mortality data were obtained from the published literature and CDC life tables, respectively. Test characteristics of cystoscopy (sensitivity and specificity to detect bladder cancer) were obtained from a large systematic review and meta-analysis of published studies (Blick et al. 2012).

Utilities associated with a given disease state, risk group, or procedure were obtained from the published literature. Costs of cystoscopy and cystectomy were obtained from Aim 2 using the Medicare claims data and the published literature.

Base-case Cost-effectiveness Analysis

In order to evaluate costs per outcome achieved, we ranked the three strategies from least to most expensive and compared them sequentially. We calculated incremental differences in total costs, QALYs, total deaths, total detected recurrent cases, total detected progressed cases, and total FP cases. We then calculated incremental cost-effectiveness ratios (ICERs) for each strategy as the additional cost divided by the change in outcome (QALY gained, death averted recurrence detected, progression averted, FP averted) compared with the next less expensive alternative, removing any dominated strategies from the next sequential comparison

(Drummond, Stoddard, and Torrance 2005). Costs and outcomes were discounted at an annual rate of 3%. We used the commonly accepted in the United States threshold of $50,000 per QALY gained and a higher threshold of $100,000 per QALY gained to evaluate the cost-effectiveness of the three surveillance strategies (Neumann, Cohen, and Weinstein 2014).

Sensitivity Analyses

We performed probabilistic sensitivity analysis with 1,000 Monte Carlo simulations to show a range of plausible values of ICERs per QALY gained, given uncertainty in the input parameters, and graphed these simulations as incremental cost-effectiveness (ICER) planes. ICER planes represent the effectiveness difference per patient on the horizontal axis plotted against the difference in costs per patient on the vertical axis (Briggs, Sculpher, and Claxton 2006). We used beta distribution for binomial data, including utilities and test characteristics (sensitivity and specificity), Dirichlet distribution for multinomial data, and gamma distribution for costs (Briggs, Sculpher, and Claxton 2006). Input parameters for which ranges or 95% CI were not reported in the literature were varied by +/-10%, using a uniform distribution (Table **3**).

We also evaluated uncertainty using cost-effectiveness acceptability curves (CEACs). The CEACs allow decision makers to compare the probability of a strategy being cost-effective under a range of different WTP thresholds. We constructed acceptability curves by plotting the probability that the estimated cost-effectiveness ratio (additional cost per QALY gained) for each surveillance strategy falls below specified values of willingness to pay (WTP) using the net-benefit framework (Briggs, Sculpher, and Claxton 2006). The net-benefit framework relies on the net monetary benefit or net health benefit statistics to overcome some of the problems associated with ICERs such as having ICERs of the same sign in opposite quadrants of the cost-effectiveness plane (Briggs, Sculpher, and Claxton 2006).

Reference Cases and Perspectives

Following the Second Panel on Cost-Effectiveness recommendations (Sanders et al. 2016), we reported a base-case analysis based on the health care sector perspective. We also considered another reference case analysis based on the societal perspective, including costs related to the informal health care sector (e.g., patient-time costs, unpaid caregiver-time costs, transportation costs) and productivity-related costs (e.g., labor market earnings lost, cost of unpaid productivity due to illness, cost of uncompensated household production) (Sanders et al. 2016). Cost components were included on a best effort basis contingent on availability of data and published literature. Considering the high average age at diagnosis of 73 years for bladder cancer patients and the relatively low labor force participation rate for older adults in the US (in 2014, 18.9% for ages 70-74 and 8% for ages 75 and older) (Toossi 2015), indirect costs associated with productivity or labor market earnings loss were less of a concern in this patient population.

CHAPTER 4: FACTORS ASSOCIATED WITH RECEIPT OF SURVEILLANCE AMONG MEDICARE PATIENTS WITH NON-MUSCLE-INVASIVE BLADDER CANCER

Overview

Purpose: With a median age at diagnosis of 73 years, bladder cancer patients often live with a high disease and comorbidity burden. This study examined how clinical, demographic, and age-related factors such as functional status and comorbidities impact surveillance among older adults with non-muscle invasive bladder cancer (NMIBC).

Methods: Using SEER-Medicare data, we identified patients aged ≥66 years diagnosed with urothelial NMIBC from 2001 to 2012. We used multilevel logistic and Poisson regression to examine frequency of surveillance, defined as ≥7 (recommended), ≥4 cystoscopies (low-intensity), and number of cystoscopies during the first two years post-diagnosis. We assessed functional status in the 12 months prior to diagnosis using Davidoff's Disability Status (DS) measure. Comorbidities were measured using the Klabunde adaptation of the Charlson comorbidity index modified for cancer patients.

Results: Among the 53,385 patients analyzed, 34% received ≥7 cystoscopies during the first two years post-diagnosis (mean=5, SD=2.9). In fully adjusted models, patient characteristics associated with the lowest odds of receiving ≥7 cystoscopies were: age ≥85 at diagnosis compared with 66-69 (adjusted odds ratio [aOR] 0.53, 95% confidence interval [95%CI] 0.49-0.57), poor DS at baseline compared with good DS (aOR 0.53, 95%CI 0.47-0.59), and having ≥3 comorbid conditions compared with no comorbidities (aOR 0.66, 95%CI 0.60-0.72). Even lower ORs were observed for the receipt of ≥4 cystoscopies.

Conclusion: Age, poor functional status, and higher comorbidity were negatively associated with surveillance frequency. Professional societies recommending risk-stratified surveillance might also consider including age, disability status, and high comorbidity when determining frequency of NMIBC surveillance.

Keywords: non-muscle-invasive bladder cancer, surveillance, functional status, comorbidities, SEER-Medicare

Background

Bladder cancer is the sixth most common cancer in the United States, and its incidence has increased from ~50,000 new cases diagnosed in 1990 to more than 80,000 new cases expected in 2018 (Silverberg, Boring, and Squires 1990; Siegel, Miller, and Jemal 2018). Over the same time period, the median age at diagnosis has increased from 65 to 73 years, the highest age among all cancer sites (American Cancer Society 2016). Most incident cases (70%) are non-muscle-invasive bladder cancer (NMIBC) at the time of presentation and a growing subgroup of predominantly older patients have lower-risk disease (Nielsen et al. 2014). Low-grade NMIBC constitutes about two thirds of all NMIBC cases; in this group approximately 50% recur, and <5% progress to muscle-invasive disease (Sylvester et al. 2006). Bladder cancer patients have one of the highest comorbidity burdens compared with other cancer patients, increasing progressively with age (Bluethmann, Mariotto, and Rowland 2016), and thus have medically complex healthcare needs (Garg, Young, et al. 2018).

Regular surveillance based on repetitive invasive endoscopy procedures is recommended for NMIBC patients of all ages to detect potential recurrences, often for the remainder of a

patient's lifespan (Chang et al. 2016; Hall et al. 2007; Power and Izawa 2016; Babjuk et al. 2016; Babjuk et al. 2013; Babjuk et al. 2011; Spiess et al. 2017). Historic guidelines for bladder cancer surveillance in the United States were consensus- rather than evidence-based and recommended cystoscopy every 3 months for 2 years post-diagnosis, then every 6 months for 2 years, and annually thereafter, regardless of patient age or underlying risk of recurrence (Hall et al. 2007). Guidelines were revised in 2016, again predominantly based on expert opinion, to recommend surveillance every 3 months for 2 years only for the minority of patients at high risk of recurrence and less frequently for those at low and intermediate risk (Chang et al. 2016; Spiess et al. 2017), similar to recommendations from other countries (Power and Izawa 2016). However, they did not explicitly consider age, multiple comorbidities, and functional impairment. Prior studies have indicated adherence to historic bladder cancer surveillance protocols is low (Schrag et al. 2003; Chamie et al. 2011; Chamie et al. 2012). It is not known whether trends in NMIBC surveillance in the United States have changed over time. Moreover, functional impairment in older bladder cancer patients, which plays an important role in health outcomes and receipt of recommended care, has not been previously investigated (Brezinski et al. 1991; Cesari et al. 2009; Cesari et al. 2008; Cesari et al. 2005; Penninx et al. 2000; Stommel, Given, and Given 2002; Inouye et al. 1998; Shih et al. 2015).

Our objectives were to examine recent trends in bladder cancer surveillance and evaluate clinical and demographic factors associated with frequency of surveillance among older US adults diagnosed with NMIBC between 2001 and 2012. We also sought to understand how age-related factors such as functional status and comorbidities impact surveillance patterns.

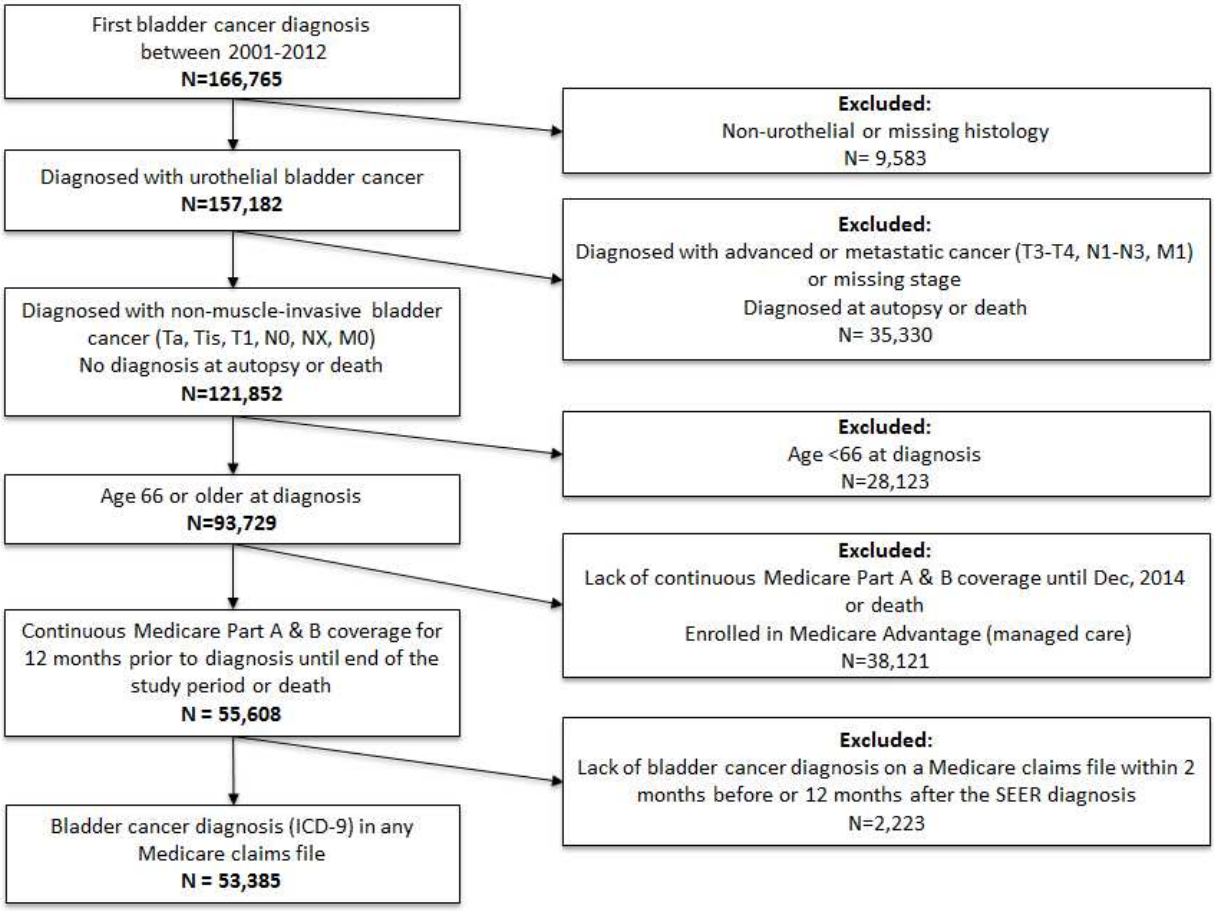
Methods

Data and Study Population

We used the Surveillance, Epidemiology and End Results (SEER) database linked with Medicare fee-for-service claims from 2000 to 2014. The SEER data contain longitudinal demographic and incident cancer characteristics for ~28% of the US cancer population (Warren et al. 2002). Medicare provides health insurance for 97% of Americans aged 65 and older, and claims reflect longitudinal healthcare encounters that capture data on diagnoses and procedures (Warren et al. 2002).

The study population (Figure 4) included SEER-Medicare beneficiaries diagnosed with their first bladder cancer between January 1, 2001 and December 31, 2012 (SEER cancer codes C67.0-C67.9; International Classification of Diseases, Ninth Revision [ICD-9] codes 188.0-188.9, 233.7). Next, we identified patients with urothelial (histology codes 8050-8052, 8120-8124, 8130-8131) non-muscle invasive (Ta, Tis, T1, N0, NX, M0) disease, excluding patients diagnosed at autopsy or death. Patients aged 66 years and older were included in the study to have one year of claims prior to diagnosis to assess comorbidities and disability status (Klabunde, Warren, and Legler 2002; Klabunde et al. 2000; Davidoff et al. 2013). Patients had to be continuously enrolled in Medicare parts A and B for a minimum of 12 months prior to first recorded diagnosis until the end of the study period or death. Patients were excluded if they were enrolled in Medicare Advantage (from the year before diagnosis through the end of the study period or death because individual claims for those enrollees are not available (Warren et al. 2002). Patients were also excluded if they did not have a confirmed bladder cancer diagnosis (ICD-9) in any Medicare claims file within 2 months before or up until 12 months after the SEER diagnosis. The final cohort included 53,385 patients.

Figure 4. Flowchart of the study cohort based on the inclusion/exclusion criteria



Outcomes

Cystoscopy was identified using Current Procedural Terminology (CPT) codes on outpatient, inpatient, carrier (physician/supplier), and durable medical equipment claims files using previously established code sets from inpatient files (Appendix 3.2). Receipt of recommended surveillance cystoscopy was operationalized as a binary variable indicating receipt of $\geq$ 7 cystoscopies over the first two years after diagnosis based on the US historic recommendations of cystoscopy evaluation every 3-6 months for all patients during the study period (Hall et al. 2007). We also examined the number of cystoscopies received over the first two years after diagnosis as a count variable.

Covariates

Covariates obtained from the Patient Entitlement and Diagnosis Summary File (PEDSF) included age at diagnosis, race/ethnicity, marital status at diagnosis, tumor grade, T classification (Ta, Tis, T1) based on the American Joint Committee on Cancer (AJCC) TNM staging system (Edge et al. 2010), prior history of cancer, year of diagnosis, history of state buy-in in the year prior diagnosis (proxy for whether a beneficiary was enrolled in a state-administered Medicaid program), SEER region, and extent of urbanization at patients' residence. We assessed patients' zip code-level socioeconomic information by using US Census data to derive quartiles of median household income in subject's zip code and proportion of adult residents with less than high school education in subject's zip code. Comorbidities were measured using the Klabunde adaptation of the Charlson comorbidity index modified for cancer patients using patients' Medicare Part A and B claims during the 12 months before diagnosis (Klabunde et al. 2007; Klabunde et al. 2000). Surgeon volume was defined as the number of transurethral resections performed by each patients' treating physician in the year prior to patient's diagnosis.

We assessed functional status in the 12 months prior to cancer diagnosis using the Disability Status (DS) measure (Davidoff et al. 2013; Davidoff 2014) which was developed using data from the Medicare Current Beneficiary Survey (MCBS) linked to Medicare Parts A and B claims (Davidoff et al. 2013). Responses to the functional status measures on the MCBS were used to construct a proxy measure for the 6-level Eastern Cooperative Oncology Group (ECOG) Performance Status (PS) scale; this measure was subsequently collapsed into a dichotomous indicator for good (0-2) versus poor (3-4) DS (5 on the ECOG PS scale indicates death) (Davidoff et al. 2013). Davidoff et al. developed a claims-based prediction algorithm including various indicators for healthcare services in the 12 months prior to cancer diagnosis

that were expected to vary based on DS level (Davidoff et al. 2013). Age and chronic conditions were not included in the DS prediction model as DS was intended to capture a health status dimension independent of these factors, consistent with the approach used in the ECOG PS scale (Davidoff et al. 2013). Following Davidoff's study validating the DS measure in four cohorts of cancer patients (Davidoff 2014), we generated the dichotomous indicator of good/poor predicted DS from the prediction model with no interactions using a cut-point of 0.110 which maximized sensitivity and specificity (Davidoff et al. 2013; Davidoff 2014). In sensitivity analysis, we also examined results for predicted DS measure based on the model with interactions using a cut-point of 0.115, but similar to Davidoff et al., we did not find meaningful differences in the results between the models with and without interactions (Davidoff 2014).

We identified the physician who performed the initial cystoscopy for each patient using the Unique Physician Identification Number (UPIN) or (National Provider Identifier) NPI included in the Medicare claims files. When a patient had more than two procedures, we assigned that patient to the provider who performed the greatest number of procedures. Following prior studies' method for assigning bladder cancer patients to providers (Schrag et al. 2003), if a patient had two urologic procedures performed by two different providers, we selected the provider who performed the second procedure, under the assumption that the provider who had the most recent contact with the patient is likely to exert the greatest influence on that patient's surveillance behavior.

Statistical Analysis

Descriptive statistics were calculated for each variable as percentages for categorical variables and means/standard deviations for continuous variables. We compared baseline characteristics between patients with poor and good DS using Chi-square test for categorical

variables and two-sided Student's t test for continuous variables. To examine factors associated with receipt of surveillance, we used two-level logistic regression models for our binary outcomes and multilevel Poisson regression for the count outcome (P. C. Austin et al. 2017). In sensitivity analysis, we considered low-intensity surveillance, defined as 4 or more cystoscopies over the first two years post-diagnosis, similar to prior studies (Chamie et al. 2011; Chamie et al. 2012). All multivariable models accounted for clustering of patients within physicians and allowed the intercept term to vary randomly over physician groups (T. A. B. Snijders and Bosker 2012; Curran and Bauer 2007). The models also included an offset term to account for the number of days each patient was observed and "at risk" for surveillance until death or total cystectomy. Patients who underwent total cystectomy were censored, as they were no longer eligible for surveillance as they had their bladder removed.

We estimated the provider-attributable residual intraclass correlation coefficient (ICC), representing unexplained physician-level variance (T. A. B. Snijders and Bosker 2012), from the full models for each binary outcome measure.

Results

We identified 53,385 patients diagnosed with NMIBC between 2001-2012 who were nested within 5,515 providers. The mean number of patients per provider was 9 (range: 1-101). The majority of patients were non-Hispanic white (92.9%), male (75.4%), had no record of comorbid conditions within the year prior to diagnosis (70.5%), were married (60.6%), and had Ta tumors (62.1%) (Table **10**). The mean age at diagnosis was 77.8 years (SD=7.0), with more than 40% of the population being older than 80 years. While only 6.4% of patients were classified as having poor DS at diagnosis, there were notable differences in patient characteristics between patients with poor vs. good DS (Table **10**). Patients with poor DS at baseline were older

at diagnosis compared with those with good DS (81.4 vs. 77.6 years, P<0.001). A higher proportion of patients with poor DS had ≥3 comorbidities at diagnosis compared with those with good DS (30.3% vs. 5.7%, P<0.001). Compared with patients with good DS, a higher proportion of those with poor DS were non-Hispanic black (3.0% vs. 8.0%, respectively) and had a history of any state buy-in coverage in the 12 months prior to diagnosis, indicating potential Medicaid dual eligibility (6.7% vs. 37.3%, respectively).

Table 10. Baseline characteristics of the study population for the full sample and stratified by predicted disability status (DS) at diagnosis

	Full Sample	**Good DS**	**Poor DS**	
Variable	**N (%)**	**N (%)**	**N (%)**	**P value**
Total number of patients	53,385 (100)	49,945 (93.56)	3,440 (6.44)	
Mean age at cancer diagnosis (SD)	77.82 (7.01)	77.57 (6.90)	81.43 (7.61)	<0.0001
Age group at cancer diagnosis				<0.0001
66-69	7,248 (13.58)	6,994 (14.00)	254 (7.38)	
70-74	11,575 (21.68)	11,118 (22.26)	457 (13.28)	
75-79	13,088 (24.52)	12,451 (24.93)	637 (18.52)	
80-84	11,630 (21.79)	10,798 (21.62)	832 (24.19)	
≥85	9,844 (18.44)	8584 (17.19)	1260 (36.63)	
Sex				<0.0001
Male	40,226 (75.35)	38,126 (76.34)	2,100 (61.05)	
Female	13,159 (24.65)	11,819 (23.66)	1,340 (38.95)	
Race/ethnicity				<0.0001
White (Non-Hispanic)	49,567 (92.85)	46,609 (93.32)	2,958 (85.99)	
Black (Non-Hispanic)	1,792 (3.36)	1,517 (3.04)	275 (7.99)	
Other (includes Other, Asian, Hispanic, and North American Native)	2,026 (3.80)	1,819 (3.64)	207 (6.02)	
Marital status at diagnosis				<0.0001
Married/domestic partner [a]	32,331 (60.56)	2,978 (5.96)	363 (10.55)	
Other (separated, divorced, widowed)	14,107 (26.43)	31,103 (62.27)	1,228 (35.70)	
Single (never married)	3,341 (6.26)	12,469 (24.97)	1,638 (47.62)	
Unknown	3,606 (6.75)	3,395 (6.80)	211 (6.13)	
Charlson comorbidity index				
0	37,633 (70.49)	36,498 (73.08)	1,135 (32.99)	<0.0001
1	8,239 (15.43)	7,544 (15.10)	695 (20.20)	
2	3,626 (6.79)	3,058 (6.12)	568 (16.51)	

≥3	3,887 (7.28)	2,845 (5.70)	1,042 (30.29)	
Predicted Disability Status (DS)[b]				
Good DS (0-2)	49,945 (93.56)			
Poor DS (3-4)	3,440 (6.44)			
Prior history of cancer	15,604 (29.23)	14,577 (29.19)	1,027 (29.85)	0.4043
Secondary cancers after bladder cancer diagnosis	8,477 (15.88)	8,187 (16.39)	290 (8.43)	<0.0001
T classification				<0.0001
Ta	33,147 (62.09)	31,208 (62.48)	1,939 (56.37)	
Tis	3,996 (7.49)	3,774 (7.56)	222 (6.45)	
T1	16,242 (30.42)	14,963 (29.96)	1,279 (37.18)	
Tumor grade				<0.0001
G1 (well differentiated)	8,475 (15.88)	7,979 (15.98)	496 (14.42)	
G2 (moderately differentiated)	17,298 (32.40)	16,251 (32.54)	1,047 (30.44)	
G3 (poorly differentiated)	9,348 (17.51)	8,681 (17.38)	667 (19.39)	
G4 (undifferentiated)	9,422 (17.65)	8,741 (17.50)	681 (19.80)	
Unknown	8,842 (16.56)	8,293 (16.60)	549 (15.96)	
Any Medicaid state buy-in in the year prior to diagnosis	4,642 (8.70)	3,358 (6.72)	1,284 (37.33)	<0.0001
SEER region				0.0015
Northeast	14,121 (26.45)	13,154 (26.34)	967 (28.11)	
Midwest	6,959 (13.04)	6,462 (12.94)	497 (14.45)	
South	11,507 (21.55)	10,792 (21.61)	715 (20.78)	
West	20,798 (38.96)	19,537 (39.12)	1,261 (36.66)	

[a] Includes having a domestic partner (same sex or opposite sex or unregistered)
[b] Dichotomous indicator of good/poor predicted DS generated from the prediction model with no interactions using a cut-point of 0.110 (Davidoff et al. 2013).

A third of patients received ≥7 cystoscopies over the first two years post-diagnosis and the baseline patient characteristics differed significantly between patients who received ≥7 vs. <7 (Appendix 4.2). The mean number of cystoscopies in the first two years post-diagnosis was 5 (SD=2.9). In the fully adjusted models (Figure **5**), patient characteristics associated with the lowest odds of receiving ≥7 cystoscopies were: poor DS at baseline compared with good DS (OR 0.53, 95%CI 0.47-0.59), age ≥85 at diagnosis compared with 66-69 (OR 0.53, 95%CI 0.49-0.57), and having ≥3 comorbid conditions compared with no comorbidities (OR 0.66, 95%CI 0.60-

0.72). The magnitude of the decrease was even greater for patients receiving ≥4 vs. <4 cystoscopies (Figure **6**).

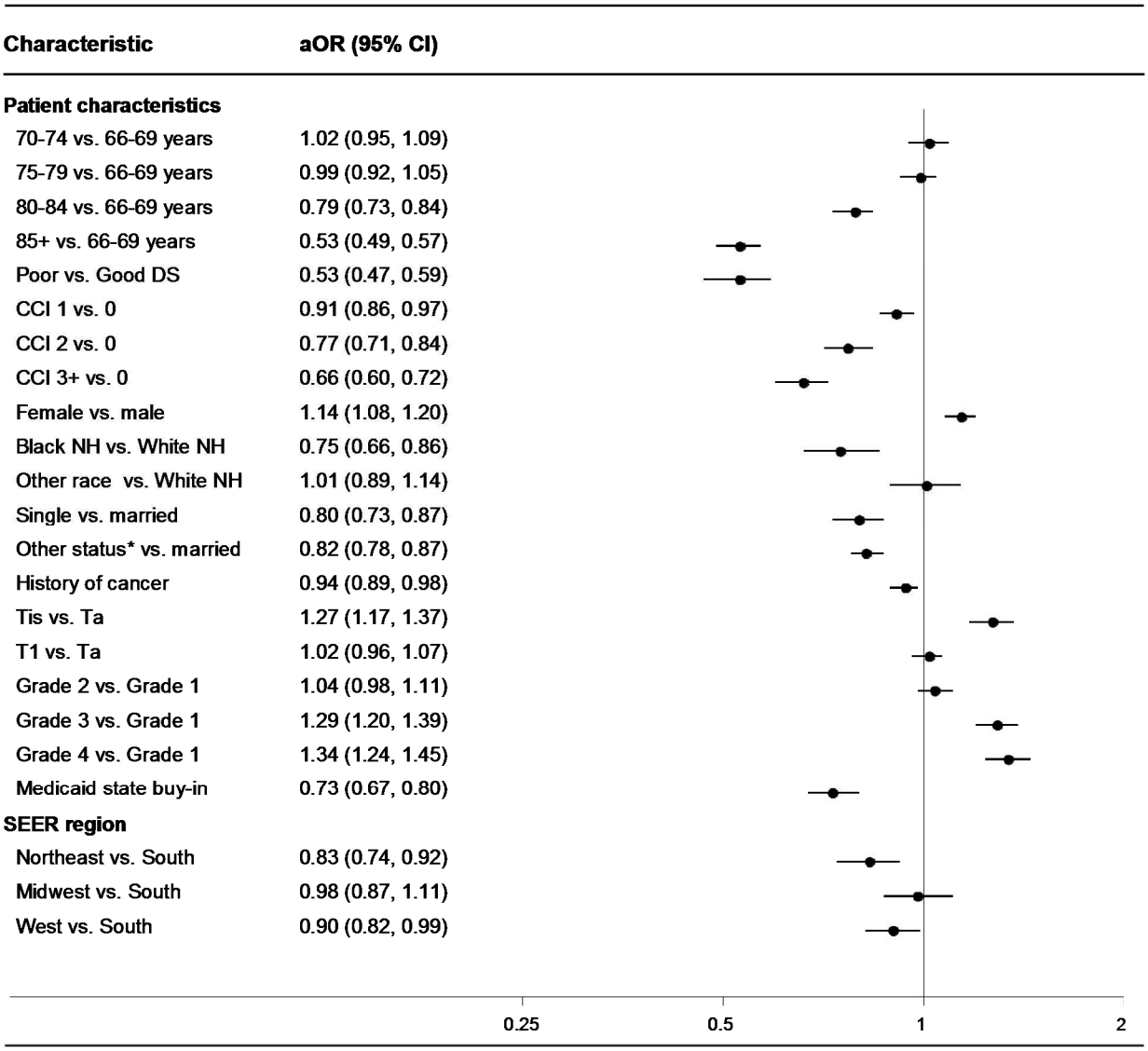
After adjusting for clinical, demographic, and socioeconomic factors, black patients were significantly less likely to receive ≥7 cystoscopies, compared with white patients (OR 0.75, 95%CI 0.66-0.86). Patients with history of any Medicaid state buy-in in the year prior to diagnosis also had significantly lower odds of receiving ≥7 cystoscopies (OR 0.73, 95%CI 0.67-0.80). Other factors associated negatively with receipt of ≥7 cystoscopies included age 80-84 at diagnosis compared with 66-69 (OR 0.79, 95%CI 0.73-0.84), being unmarried at the time of diagnosis, prior history of cancer (OR 0.94, 95%CI 0.89-0.98), having one (OR 0.91, 95%CI 0.86-0.97) or two comorbidities (OR 0.77, 95%CI 0.71-0.84) compared with no comorbidities. These factors were associated with even lower odds of receiving ≥4 vs. <4 cystoscopies. Factors associated positively with receipt of ≥7 cystoscopies were being female (OR 1.14, 95%CI 1.08-1.20), Tis classification (OR 1.27, 95%CI 1.17-1.37), and being diagnosed with poorly differentiated (OR 1.29, 95%CI 1.20-1.39) or undifferentiated tumor (OR 1.34, 95%CI 1.24-1.45).

16% of the variance of receipt of ≥7 cystoscopies was explained by the between-provider variability (ICC 0.160, 95%CI 0.148-0.173). Only 7.8% of the variance of receipt of ≥4 cystoscopies was explained by the between-provider variability (ICC 0.078, 95%CI 0.069-0.089).

The multivariable two-level Poisson model yielded similar results for factors associated with receipt of number of cystoscopies (Figure **7**). The three factors associated with the lowest rate of cystoscopy over the first two years after diagnosis were: age ≥85 compared with 66-69 (IRR 0.88, 95%CI 0.87-0.90), poor DS at baseline compared with good DS (IRR 0.89, 95%CI

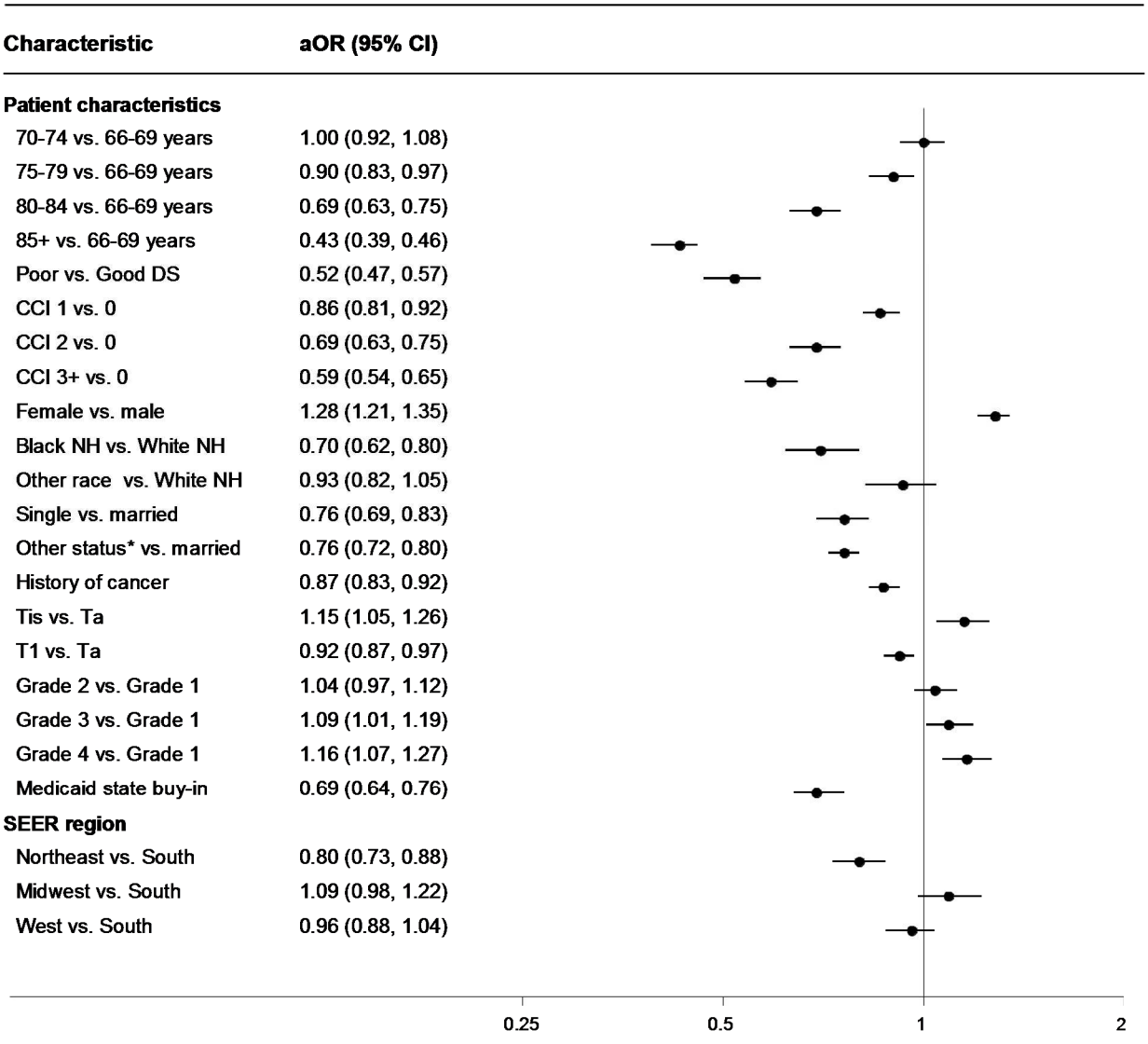
0.87-0.91), and being black compared with white (IRR 0.92, 95%CI 0.90-0.94). Having ≥3 comorbid conditions was associated with a small decrease in the rate of cystoscopy compared with having no comorbidities (IRR 0.96, 95%CI 0.95-0.98).

Figure 5. Results from the fully-adjusted logistic regression model of factors associated with receipt of ≥7 cystoscopies



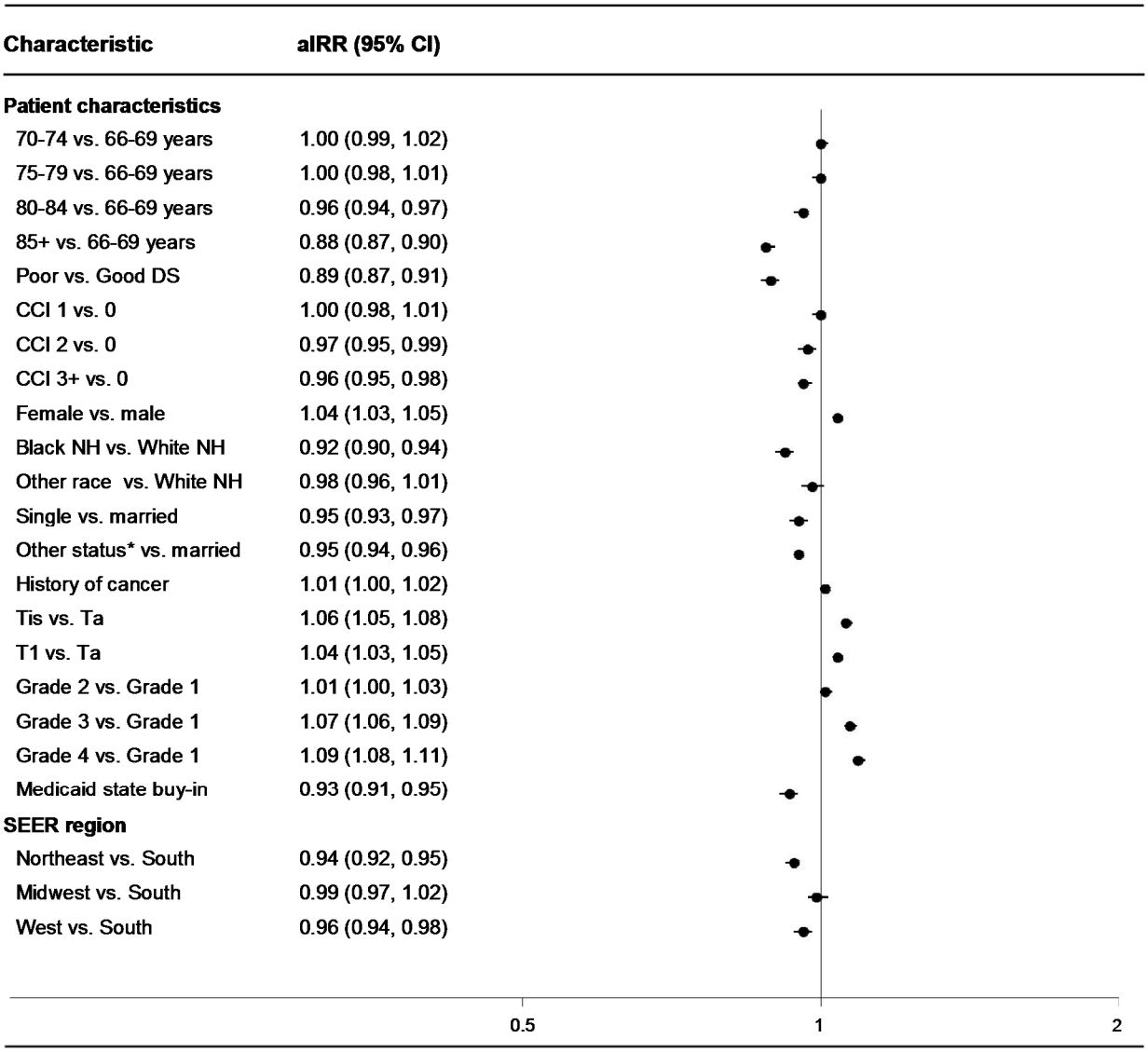
Note: The model was also adjusted for zip-code level education and median household income, residential status at diagnosis, surgeon volume, and year of cancer diagnosis.
Abbreviation: aOR, adjusted odds ratio; CI, confidence interval; NH, non-Hispanic.
* Other status includes separated, divorced, or widowed at diagnosis.

Figure 6. Results from the fully-adjusted logistic regression model of factors associated with receipt of ≥4 cystoscopies



Note: The model was also adjusted for zip-code level education and median household income, residential status at diagnosis, surgeon volume, and year of cancer diagnosis.
Abbreviation: aOR, adjusted odds ratio; CI, confidence interval; NH, non-Hispanic.
* Other status includes separated, divorced, or widowed at diagnosis.

Figure 7. Results from the fully-adjusted Poisson regression model of factors associated with number of cystoscopies received



Note: The model was also adjusted for zip-code level education and median household income, residential status at diagnosis, surgeon volume, and year of cancer diagnosis.
Abbreviation: aIRR, adjusted incidence rate ratio; CI, confidence interval; NH, non-Hispanic.
* Other status includes separated, divorced, or widowed at diagnosis.

Discussion

Consideration of comorbidities and functional status is important for determining optimal care among older bladder cancer patients. The three factors associated with the lowest odds of receipt of recommended surveillance were older age, poor disability status, and increased

comorbidities. Similar to other studies, we found that older age at diagnosis was inversely associated with frequency of surveillance (Schrag et al. 2003; Chamie et al. 2011). Notably, in the fully adjusted models, age ≥ 85 at diagnosis was associated with the lowest odds of receiving ≥7 and ≥4 cystoscopies, followed by poor DS at diagnosis.

After adjusting for clinical, demographic, and socioeconomic factors, black patients compared with white patients were significantly less likely to receive cystoscopy across all three surveillance outcomes. At the same time, a higher proportion of patients with poor DS were black compared with those with good DS (8.0% vs. 3.0%), suggesting higher burden of functional decline among black patients. History of any Medicaid state buy-in in the year prior to diagnosis also had significant negative impact on surveillance use and was related to race: a higher proportion of patients with any state buy-in history were black compared with those with no buy-in (9.9% vs. 2.7%).

We found greater surveillance among Medicare patients with NMIBC than previously reported (Schrag et al. 2003; Chamie et al. 2011; Chamie et al. 2012): 34% of patients in our cohort received ≥7 and 70% received ≥4 cystoscopies over the first two years post-diagnosis. By contrast, only 40% of patients with superficial bladder cancer diagnosed between 1992 and 1996 underwent surveillance once every six months for three years (Schrag et al. 2003). Chamie et al. found even lower rates of adherence to historic surveillance cystoscopy recommendations: among high-grade NMIBC patients diagnosed between 1992-2002, only 4.9% received eight or more cystoscopies and 33.6% received at least four cystoscopies over the first two years after diagnosis (Chamie et al. 2011). A possible reason for the higher surveillance use in our study is that we examined a broader cohort of NMIBC patients diagnosed more recently (2001-2012) and included patients with both high and low grade tumors, prior history of cancer, and secondary

cancer diagnoses. Unlike prior reports, we did not exclude patients who died or underwent total cystectomy during the exposure period (Schrag et al. 2003; Chamie et al. 2011; Chamie et al. 2012). Instead, we accounted for the different exposure time and patient-days during which each subject was observed and "at risk" for receiving surveillance. Moreover, all previous studies on bladder cancer surveillance lacked important information about patients' functional status, which may influence patients' and providers' decisions to undertake regular surveillance cystoscopy (Schrag et al. 2003; Chamie et al. 2011; Chamie et al. 2012; Hollingsworth et al. 2010).

Given that overuse of surveillance among low-risk, and underuse among high-risk, patients may have undesirable consequences (Schroeck, Smith, and Shelton 2018), there has been increased impetus towards risk-aligned bladder cancer surveillance. The revised 2016-2017 guidelines in the United States now recommend surveillance every 3 months for 2 years only for the minority of patients at high risk of bladder cancer recurrence and less frequently for those at low and intermediate risk (Chang et al. 2016; Spiess et al. 2017). However, age, multiple comorbidities, and functional impairment have not yet been explicitly considered by current bladder cancer guidelines. The American Society of Clinical Oncology has underscored the importance of incorporating multiple chronic conditions in guideline development and implementation, cautioning that guideline recommendations that neglect this issue may not apply to, or could be harmful for, patients with multiple chronic conditions (Somerfield et al. 2016). National Comprehensive Cancer Network guidelines for prostate cancer already consider life expectancy, comorbidities, and age in determining appropriate intensity of care (Mohler et al. 2016; Carroll et al. 2016). Compared with prostate cancer patients, bladder cancer patients had higher comorbidity burden, increasing progressively with age (Bluethmann, Mariotto, and Rowland 2016). Patients with urological cancer also had disproportionate burden of multiple

comorbid conditions and medical complexity, compared with primary care patients, highlighting the need to incorporate comorbidities in personalized treatment and surveillance recommendations (Garg, Young, et al. 2018).

Limitations of this study should also be considered. SEER-Medicare data do not collect follow-up information on patients' tumor recurrence or progression; neither do they capture underlying risks of recurrence or progression which may impact surveillance decisions. Furthermore, we could not account for patient and provider preferences for surveillance.

Conclusion

We found that NMIBC patients aged ≥85 years and those with poor disability status and ≥3 comorbidities were least likely to undergo frequent surveillance. As the age at diagnosis and number of comorbid conditions increased, the odds of receiving recommended cystoscopy frequency as well as the rate of cystoscopy decreased. Professional societies recommending risk-stratified surveillance based on tumor factors might also consider including age, disability status, and high comorbidity when determining appropriate frequency of surveillance for NMIBC patients.

CHAPTER 5: SURVIVAL PATTERNS AMONG MEDICARE PATIENTS WITH NON-MUSCLE-INVASIVE BLADDER CANCER UNDERGOING SURVEILLANCE

Overview

Purpose: Non-muscle-invasive bladder cancer (NMIBC) is a heterogeneous disease characterized by variable risks of recurrence, progression, and death. Comorbidities also represent important cause of mortality in NMIBC patients. The purpose of this study was to characterize survival patterns among older patients with NMIBC undergoing surveillance cystoscopy and examine differences in bladder-cancer-specific and other-cause mortality by age, comorbidities, and functional status.

Methods: In a retrospective analysis of SEER-Medicare data, we identified patients aged ≥66 years diagnosed with urothelial NMIBC from 2001 to 2012. We used propensity-score weighted Fine-Gray competing risk regression and Cox proportional hazards model to assess all-cause, bladder-cancer-specific and other-cause mortality after receipt of high-intensity (≥7 vs. <7 cystoscopies) or low-intensity surveillance (≥4 vs. <4 cystoscopies) during the first two years after diagnosis.

Results: In the analyzed cohort of 41,743 patients with NMIBC, both receipt of high-intensity and low-intensity surveillance were associated with a decrease in the hazard of all-cause mortality in the propensity-score weighted survival models. Receipt of ≥7 cystoscopies was associated with an increase in the hazard of bladder-cancer death and a decrease in the hazard of other-cause death, which could be due to potential unmeasured confounding by indication. Older patients (≥75 vs. 66-74 years) and those with poor disability status at diagnosis

had higher cumulative incidence of both bladder-cancer and other-cause death, regardless of frequency of cystoscopy.

Conclusion: Higher bladder-cancer mortality in older patients and those with poor disability status, regardless of frequency of cystoscopy, warrants further research. Randomized controlled trials and accounting for time-varying confounding are needed to assess the survival benefits and potential harms of different frequencies of surveillance among older patients with NMIBC.

Keywords: non-muscle-invasive bladder cancer, surveillance, competing risks, mortality, SEER-Medicare

Background

Bladder cancer imposes a significant burden of disease in the United States with 2018 estimates predicting 81,190 incident cases diagnosed and more than 17,000 deaths (Siegel, Miller, and Jemal 2018). Approximately 3 in 4 newly diagnosed cases are non-muscle invasive bladder cancer (NMIBC). NMIBC is a heterogeneous disease characterized by variable risks of recurrence, progression, and death, depending on tumor characteristics. Roughly 50-70% of NMIBC will recur and 10–20% will progress to muscle-invasive disease (MIBC) (Sylvester et al. 2006; Fernandez-Gomez et al. 2009). Of all newly diagnosed NMIBC, 60-70% present as noninvasive papillary carcinoma (Ta), 20-30% as lamina propria-invasive disease (T1), and ~10% as carcinoma in situ (Tis) (Nielsen et al. 2014; Sylvester et al. 2006). Long-term follow-up has shown that the progression rate varies by tumor grade and T stage and is as low as 6% for low-grade Ta and as high as 17% for high-grade T1 (Palou et al. 2012; Leblanc et al. 1999). The survival prognosis for patients with NMIBC is relatively favorable, with cancer-specific survival

at 5 years ranging from 98.5% in patients with low-grade Ta to 88.7% in patients with high-grade T1 disease (Cambier et al. 2016). However, cancer-specific survival after progression from high-risk (e.g., high-grade T1) NMIBC to MIBC have been found to be much lower: 32% (range: 13%-64%) in 7 prospective trials with a median follow-up of 52-123 months and 37% (range: 7%-59%) in 12 retrospective trials with a median follow-up of 48-107 months (Van Den Bosch and Witjes 2011).

Comorbidities and comorbidity-associated events also represent important causes of mortality in bladder cancer patients, who have one of the highest comorbidity burdens compared with other cancer patients (Bluethmann, Mariotto, and Rowland 2016). Among patients newly diagnosed with bladder cancer, the 5-year cancer-specific mortality rate was found to vary between 1% and 59%, and other-cause mortality rate between 6% and 90%, depending on the tumor type and patient age (Noon et al. 2013).

Despite the variable but generally low progression rates, NMIBC has very high rates of recurrent NMIBC (Sylvester et al. 2006). Given the high recurrence rates, clinical guidelines based on expert-opinion recommend regular surveillance cystoscopy for patients with NMIBC of all ages to detect potential recurrences or progression to MIBC. Considering the evidence of a growing subgroup of predominantly older NMIBC patients with lower-risk disease and high comorbidity burden, the aim of this study was to characterize survival patterns among Medicare patients with NMIBC undergoing cystoscopy and examine differences in bladder-cancer-specific and other-cause mortality by patient age, comorbidities, and functional status.

Methods

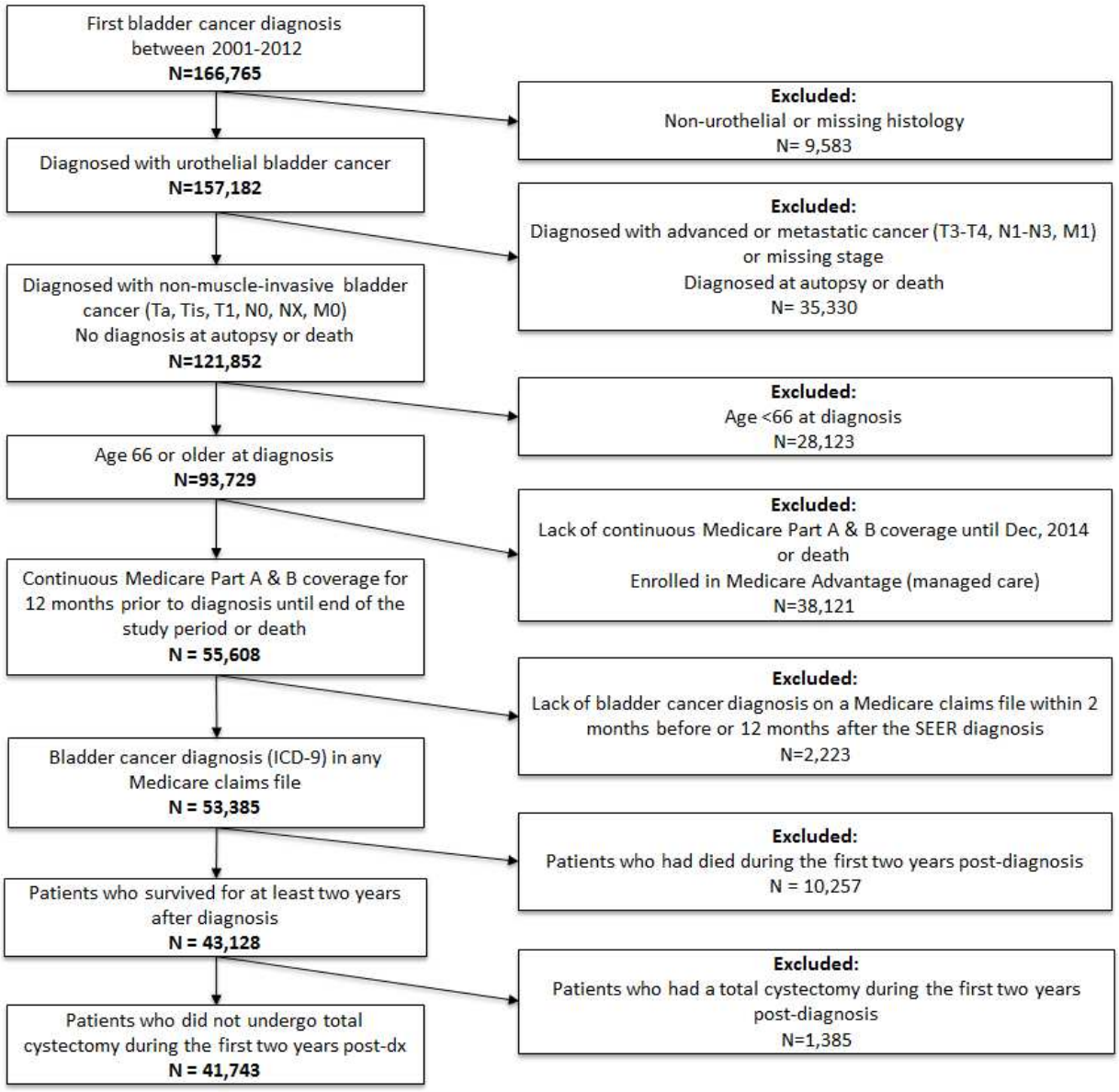
Data and Study Population

We used the Surveillance, Epidemiology and End Results (SEER) database linked with Medicare fee-for-service claims from 2000 to 2014. The SEER data contain longitudinal demographic and incident cancer characteristics for ~28% of the US cancer population (Warren et al. 2002). Medicare provides health insurance for 97% of Americans aged 65 and older, and claims reflect longitudinal healthcare encounters that capture data on diagnoses and procedures (Warren et al. 2002).

The study population included SEER-Medicare beneficiaries first diagnosed with NMIBC between January 1, 2001 and December 31, 2012 (Figure 8). We identified bladder cancer patients (SEER cancer codes C67.0-C67.9; International Classification of Diseases, Ninth Revision [ICD-9) codes 188.0-188.9, 233.7) with urothelial (histology codes 8050-8052, 8120-8124, 8130-8131) non-muscle invasive (Ta, Tis, T1, N0, NX, M0) disease. Patients aged 66 years and older were included in the study in order to have one year of claims prior to diagnosis to assess comorbidities and disability status (Klabunde, Warren, and Legler 2002; Klabunde et al. 2000; Davidoff et al. 2013). Patients had to be continuously enrolled in Medicare parts A and B for a minimum of 12 months prior to first recorded diagnosis until the end of the study period or death. Patients were excluded if they were enrolled in Medicare Advantage from the year before diagnosis through the end of the study period or death because individual claims for those enrollees are not available (Warren et al. 2002). Patients were also excluded if they were diagnosed at autopsy or death or if they did not have a confirmed bladder cancer diagnosis (ICD-9) in any Medicare claims file within 2 months before or up until 12 months after the SEER diagnosis. We additionally excluded patients who died or underwent total cystectomy during the

first two years after diagnosis as we could not assign them to a surveillance cystoscopy group, which requires two years of observed exposure. The final cohort included 41,743 patients.

Figure 8. Flowchart of the study cohorts based on the inclusion/exclusion criteria



Outcomes

The main outcomes were all-cause, bladder-cancer-specific and other-cause mortality. Mortality was determined based on the date of death, which is available in both the Medicare and

SEER data files, but derived from different sources and covers different time periods. The Medicare death date is obtained from the Medicare Enrollment Database (EDB), which is updated nightly by the Social Security Administration and is current as of the day that the enrollment data were extracted for the National Cancer Institute's use (Bach et al. 2002). Date of death information from the EDB was available through December, 2016. The SEER death date is primarily obtained from state death certificates (Bach et al. 2002) and was available from the SEER Patient Entitlement and Diagnosis Summary File (PEDSF) through December, 2013. We used the Medicare EDB file as the primary source to identify time of death for overall mortality because it allows us to have a longer observation timeframe (up to sixteen years). However, since the cause of death is only available from the SEER PEDSF, we examined cause-specific mortality only until December, 2013.

Study Variables

The primary explanatory variable – surveillance cystoscopy use – was identified using CPT codes on outpatient, inpatient, carrier (physician/supplier), and durable medical equipment (DME) claims files using previously established CPT code sets and ICD-9 codes from inpatient files (Appendix 3.2). Receipt of recommended surveillance cystoscopy was defined as a binary variable indicating receipt of 7 or more cystoscopies over the first two years after diagnosis based on the US historic recommendations of cystoscopy evaluation every 3-6 months for all patients, covering the study period (Hall et al. 2007). In sensitivity analysis, we applied another definition of low-intensity surveillance as 4 or more cystoscopies over the first two years post-diagnosis (Chamie et al. 2011; Chamie et al. 2012).

Covariates obtained from the SEER PEDSF included age at diagnosis (66-69, 70-74, 75-79, 80-84, ≥85), race/ethnicity (non-Hispanic white, non-Hispanic black, other, including

Hispanic), marital status at diagnosis (married/domestic partnership, single, other, unknown), tumor grade (G1-G4), American Joint Committee on Cancer (AJCC) T classification (Ta, Tis, T1), prior history of cancer, year of diagnosis, history of state buy-in in the year prior diagnosis (proxy for whether a beneficiary was enrolled in a state-administered Medicaid program), SEER region (Northeast, Midwest, South, West), and extent of urbanization at patients' residence.

Comorbidities were measured using the Klabunde adaptation of the Charlson comorbidity index modified for cancer patients and based on patients' Medicare Part A and B claims during the 12 months before diagnosis (Klabunde et al. 2007; Klabunde et al. 2000).

We also assessed functional status in the 12 months prior to cancer diagnosis using the Disability Status (DS) measure (Davidoff et al. 2013; Davidoff 2014). The dichotomous measure indicating poor versus good DS at baseline was constructed as a proxy for the 6-level Eastern Cooperative Oncology Group (ECOG) Performance Status (PS) scale that represents various levels of functional impairment (Davidoff et al. 2013). The DS measure was developed using a claims-based prediction algorithm including various indicators for health care services (e.g., home oxygen and respiratory therapy services, wheelchair use) in the 12 months prior cancer diagnosis that were expected to vary based on DS level (Davidoff et al. 2013). The DS measure and algorithm used to construct it are described in greater detail in Chapters 3 and 4.

We obtained patients' zip code-level socioeconomic information using US Census data to derive quartiles of median household income and proportion of adult residents with less than high school education in subject's zip code. Surgeon volume was defined as the number of transurethral resections performed by each patients' treating provider in the year prior to patient's diagnosis.

Propensity Score Estimation and Application

We estimated propensity scores by modeling the probability of receiving surveillance (≥7 cystoscopies or ≥4 cystoscopies) in the first two years following bladder cancer diagnosis as a function of all explanatory variables measured at baseline using two-level logistic regression models, accounting for clustering of patients within physicians. The second level of the two-level model was represented by the provider identification number and allowed the intercept term to vary randomly over physician groups (T. A. B. Snijders and Bosker 2012; Curran and Bauer 2007). The models adjusted for patients' age at diagnosis, sex, race/ethnicity, marital status, comorbidities and disability status at diagnosis, prior history of cancer, T classification at diagnosis, tumor grade, history of Medicaid state buy-in in the year prior diagnosis, SEER region, zip-code level education and median household income, rural/urban residence at diagnosis, surgeon volume, and year of cancer diagnosis.

We used a post-estimation function from the two-level logistic regression models to generate propensity scores and created two sets of inverse probability of treatment weights (IPTW) for each patient: a) equal to the inverse of the propensity score [1/p] for patients who received ≥7 cystoscopies and ≥4 cystoscopies; and b) equal to the inverse of one minus the propensity score [1/(1 − p)] for the two control groups (patients who received <7 and <4 cystoscopies), based on the two different surveillance definitions. We stabilized the propensity score weights by multiplying the treatment and comparison IPTW by a constant, equal to the expected value of being in the treatment or comparison groups (Robins, Hernán, and Brumback 2000; Garrido et al. 2014; Harder, Stuart, and Anthony 2011). The purpose of using stabilized weights is to reduce the weights of either those treated subjects with low propensity scores or those untreated subjects with high propensity scores (S. Xu et al. 2017). This method of

propensity score weighting provides an estimate of the treatment effect in the population (in this case, the effect of surveillance cystoscopy use among all NMIBC patients) (Stürmer, Rothman, and Glynn 2006). For before and after matching comparisons between the treatment and control groups, we examined the standardized difference in the means (SDM) which is preferred to t-tests because SDM is not influenced by sample size (P. Austin 2009; P. C. Austin 2011). If the absolute difference is <10%, the two groups are considered balanced (Normand et al. 2001).

Survival Analysis

To characterize the association between bladder cancer-specific mortality and surveillance, we used a propensity-score weighted, competing-risks regression model (Fine and Gray 1999). Failure was defined as bladder cancer mortality and the competing event as non-bladder cancer mortality. Estimates from the competing-risk models were reported as subdistribution hazard ratios (SHR) with corresponding 95% confidence intervals (95% CIs) (Fine and Gray 1999). The explanatory variable included in the propensity-score weighted survival models was an indicator for receipt of cystoscopy (≥7 vs. <7 or ≥4 vs. <4 cystoscopies over the first two years post-diagnosis).

We also used a propensity-score weighted Cox proportional hazards model to examine the association between all-cause mortality and surveillance where failure was defined as death from any cause. All survival models included standard errors clustered on the provider using the Huber-White sandwich variance estimator, because patients treated by the same provider may have similar outcomes. We confirmed the non-violation of the proportional hazards assumption (i.e., constant hazard ratio over time) using Schoenfeld's residuals test and the time-dependent covariate test for interactions between time and the cystoscopy indicators (Bradburn et al. 2003a; Bradburn et al. 2003b).

As a sensitivity analysis we estimated bladder cancer-specific and other-cause mortality using Cox proportional hazards model for cause-specific mortality where failure was defined as death from bladder cancer, while the competing risks (other-cause deaths) were treated as censored events. A major advantage of the Fine-Gray model over the Cox proportional hazards model, especially if competing risks are frequent, is that the Fine-Gray model allows for dependency between the modeled competing risks (P. C. Austin, Lee, and Fine 2016).

In sensitivity analyses, we compared results from the propensity-score weighted survival models using the stabilized inverse probability of treatment weights (SIPTW) and adjusting for the main explanatory variable (cystoscopy indicator) only 1) to those from multivariable adjusted models (including separately all covariates from the propensity score model in addition to the cystoscopy indicators) and 2) to unadjusted models (including cystoscopy indicators only).

Lastly, we generated cumulative incidence functions (CIFs) from the propensity-score weighted Fine-Gray competing risk regression to compare the cumulative incidence of bladder-cancer or other-cause death between patients who received ≥7 vs. <7 and ≥4 vs. <4 cystoscopies. We also stratified the CIFs by key covariates, including age at diagnosis, disability status, and comorbidities.

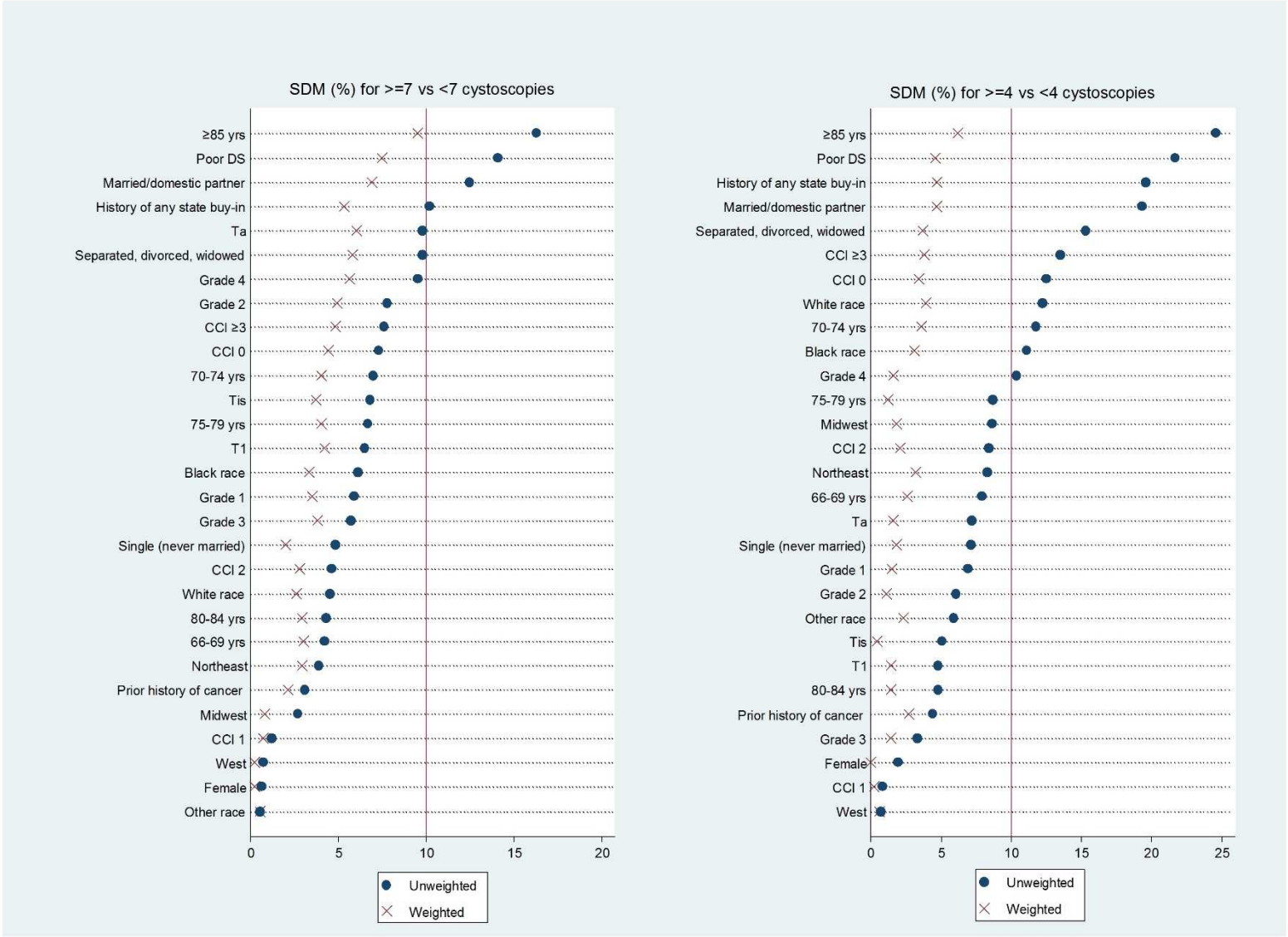
Results

We identified 41,743 patients diagnosed with NMIBC between 2001-2012 who met the study criteria (Figure **8**). The mean and median survival from diagnosis through December 2016, were 83.6 (SD 39.8) and 76 (IQR 56) months, respectively. 2,325 (5.6%) patients in the study sample died because of their bladder cancer, while 14,353 (34.4%) died from other causes. The five most common causes of death other than urinary bladder recorded from the SEER registries

were heart disease (9.7%), other cause of death (4.6%), lung and bronchus (2.8%), chronic obstructive pulmonary disease (2.8%), and cerebrovascular diseases (1.7%).

Before propensity-score weighting, SDM for the baseline characteristics between patients who received ≥7 vs. <7 cystoscopies ranged from 4% to 16%, while after applying SIPTW baseline characteristics between the two groups were well balanced, with SDM <10% for all covariates (Figure **9**). Similar balance after SIPTW was also achieved for patients who received ≥4 vs. <4 cystoscopies.

Figure 9. Standardized differences in the means (SDM) of baseline characteristics for patients who received ≥7 vs. <7 or ≥4 vs. <4 cystoscopies during the first two years post-diagnosis (before and after propensity score weighting)



Note: Additional baseline characteristics included zip-code level education and median household income, residential status at diagnosis, surgeon volume, and year of cancer diagnosis; standardized differences in the means (SDM) after propensity score adjustment were <10% for all.

Results from the survival models estimating the association between surveillance cystoscopy use and all-cause, bladder-cancer, and non-bladder-cancer death are presented in Table **11**. In the SIPTW weighted Cox proportional hazards model assessing all-cause mortality through December 2016, receipt of ≥7 vs. <7 cystoscopies over the first two years after diagnosis was associated with 8% decrease in the hazard of all-cause death (HR 0.92, 95%CI 0.90-0.94). When looking at cause-specific death (assessed through December 2013), receipt of ≥7 vs. <7 cystoscopies was associated with 15% increase in the subdistribution hazard of bladder-cancer death (SHR 1.15, 95%CI 1.05-1.17), using the Fine-Gray SIPTW weighted model, and 8% increase in the cause-specific hazard of bladder-cancer death (CHR 1.08, 95%CI 0.98-1.19), using the cause-specific Cox hazard model. The finding of an increased hazard of bladder-cancer death is likely due to residual confounding by indication—patients at high risk of recurrence and those with previous recurrences might be more likely to receive high-intensity surveillance (≥7 cystoscopies) and die due to their bladder cancer. Receipt of ≥7 vs. <7 cystoscopies was associated with 19% decrease in the subdistribution hazard of other-cause death (SHR 0.81, 95%CI 0.78-0.84), using the Fine-Gray SIPTW weighted model, and 19% decrease in the cause-specific hazard of other-cause death (CHR 0.81, 95%CI 0.78-0.84), using the SIPTW weighted cause-specific Cox hazard model.

In sensitivity analyses, lower intensity of surveillance (receipt of ≥4 vs. <4 cystoscopies) was associated with 19% reduction in the hazard of all-cause death in the SIPTW weighted model (HR 0.81, 95%CI 0.78-0.83). Receipt of ≥4 vs. <4 cystoscopies was not associated with

an increase in the hazard of bladder-cancer death. On the contrary, it was associated with 19% decrease in the subdistribution hazard of bladder-cancer death (SHR 0.81, 95%CI 0.72-0.91), using the Fine-Gray SIPTW weighted model, and 29% decrease in the cause-specific hazard of bladder-cancer death (CHR 0.71, 95%CI 0.62-0.80), using the cause-specific Cox hazard model. Receipt of ≥4 vs. <4 cystoscopies was associated with even lower hazard of other-cause death: SHR 0.69 (95%CI 0.65-0.72), using the Fine-Gray SIPTW weighted model, and CHR 0.66 (95%CI 0.63-0.70), using the cause-specific Cox hazard model.

The covariate-adjusted survival models comparing the association between receipt of ≥7 vs. <7 and ≥4 vs. <4 cystoscopies and all-cause, bladder-cancer, or other-cause death yielded similar results to those from the SIPTW weighted models (Table **11**).

Table 11. Results from the survival models on the association between surveillance cystoscopy and all-cause, bladder-cancer, and other-cause death

	Overall Cox Proportional Hazards Model	Fine-Gray Competing Risk Regression Model		Cause-Specific Cox Proportional Hazards Model	
Model	All-cause Death HR (95% CI)[a]	Bladder-cancer Death SHR (95% CI)[b]	Other-cause Death SHR (95% CI)[b]	Bladder-cancer Death CHR (95% CI)[b]	Other-cause Death CHR (95% CI)[b]
≥7 cystoscopies *(ref: <7 cystoscopies)*					
SIPTW weighted	0.92*** (0.90-0.94)	1.15*** (1.05-1.27)	0.81*** (0.78-0.84)	1.08 (0.98-1.19)	0.81*** (0.78-0.84)
Covariate adjusted	0.92*** (0.90-0.93)	1.10** (1.01-1.21)	0.81*** (0.78-0.84)	1.02 (0.93-1.11)	0.81*** (0.78-0.84)
Unadjusted	0.89*** (0.87-0.91)	1.12** (1.03-1.22)	0.75*** (0.72-0.77)	1.03 (0.95-1.13)	0.75*** (0.72-0.77)
≥4 cystoscopies *(ref: <4 cystoscopies)*					
SIPTW weighted	0.81*** (0.78-0.83)	0.81*** (0.72-0.91)	0.69*** (0.65-0.72)	0.71*** (0.62-0.80)	0.66*** (0.63-0.70)
Covariate adjusted	0.80*** (0.78-0.82)	0.84*** (0.75-0.94)	0.71*** (0.68-0.75)	0.72*** (0.64-0.80)	0.68*** (0.65-0.71)
Unadjusted	0.77*** (0.74-0.79)	0.80*** (0.72-0.89)	0.61*** (0.59-0.64)	0.67*** (0.61-0.75)	0.59*** (0.56-0.61)

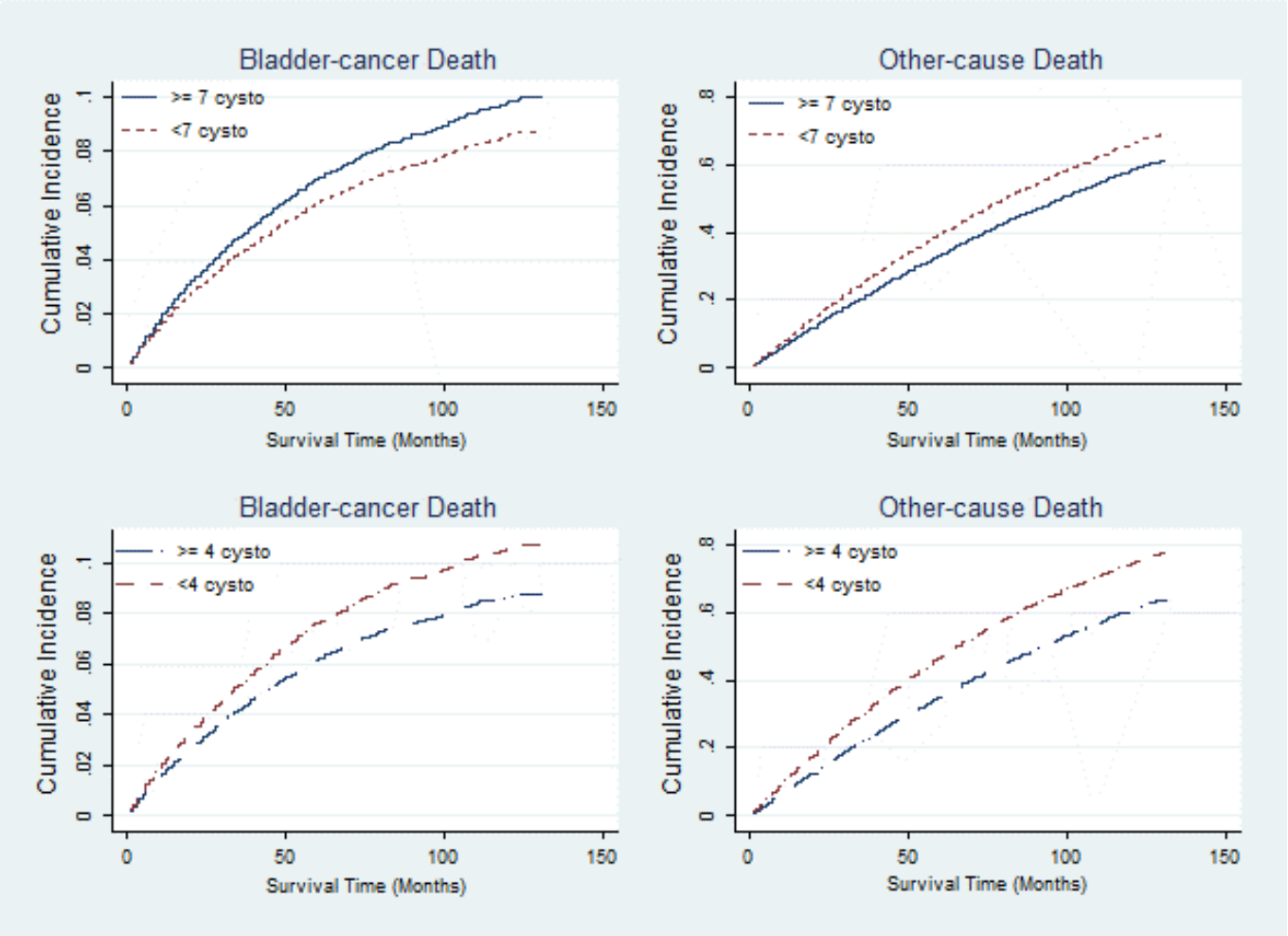
[a] Death date for all-cause mortality was assessed through December, 2016 using the Medicare EDB file.
[b] Death date for cause-specific mortality was assessed through December, 2013 using the SEER PEDSF.
Abbreviations: HR, hazard ratio; 95% CI, 95% confidence interval; SHR, subdistribution hazard ratio; CHR, cause-specific hazard ratio; ref, referent; SIPTW, stabilized inverse probability of treatment weights.
Note: *Statistically significant at 10%; **Statistically significant at 5%; ***Statistically significant at 1%.

Cumulative incidences of bladder-cancer and other-cause death generated from the propensity-score weighted Fine-Gray competing risk regression models by frequency of surveillance (≥7 vs. <7 and ≥4 vs. <4 cystoscopies) are described in Figure 10. The cumulative incidence of other-cause death exceeded that of bladder cancer death at each point in time, regardless of cystoscopy use. The cumulative incidence of bladder-cancer death was higher among patients receiving ≥7 cystoscopies compared with <7 cystoscopies. However, the cumulative incidence of other-cause death was lower among patients receiving ≥7 cystoscopies

compared with <7 cystoscopies. The cumulative incidences of both bladder-cancer death and other-cause death were higher among patients receiving <4 cystoscopies compared with ≥4 cystoscopies.

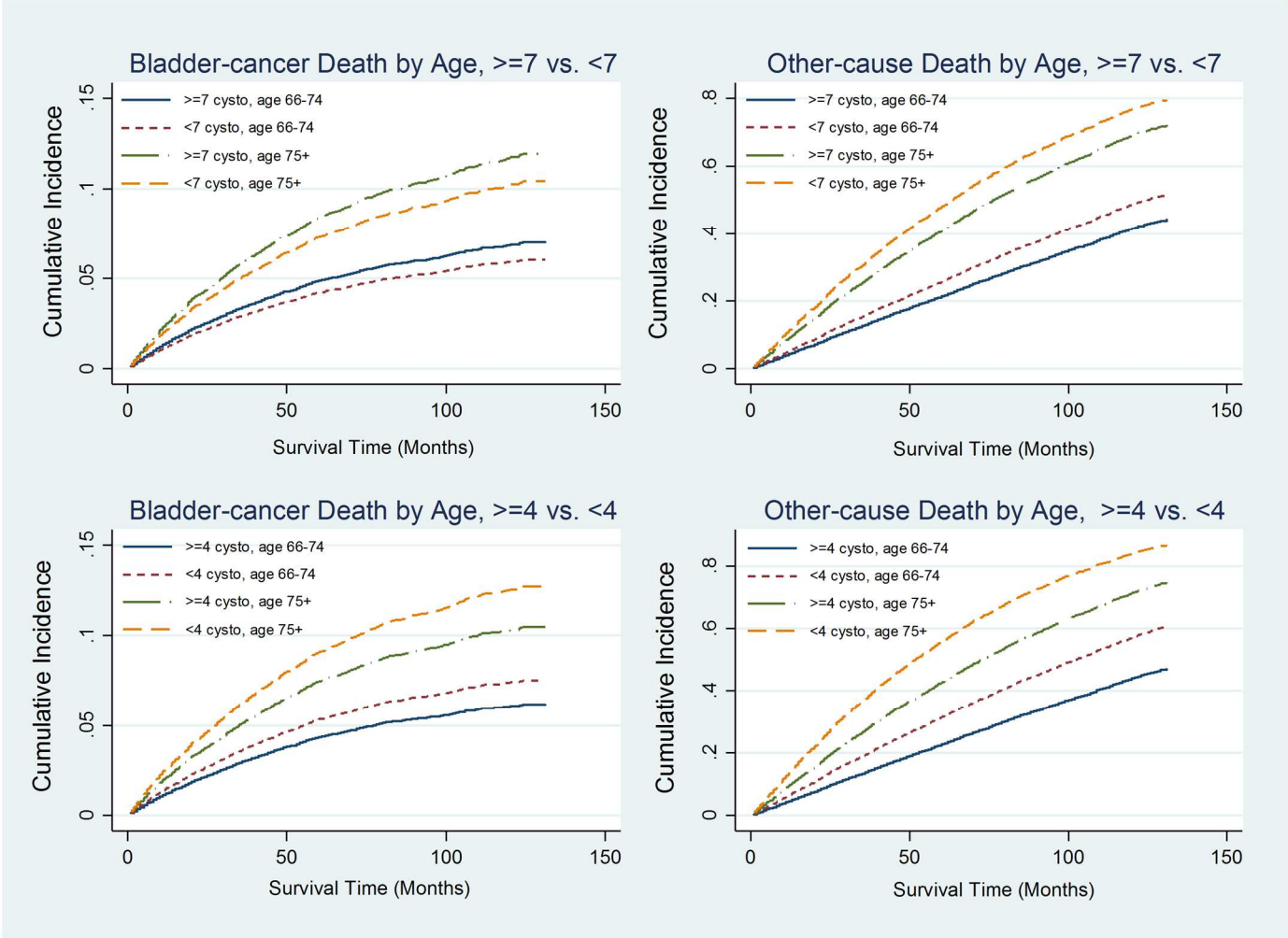
CIFs by receipt of surveillance (≥7 vs. <7 and ≥4 vs. <4 cystoscopies), generated from the propensity-score weighted Fine-Gray competing risk regression models and stratified by age at diagnosis (66-74 vs. ≥75), disability status (poor vs. good DS), and comorbidities (0 vs. ≥1), are presented in Figure **11**, Figure **12**, and Figure **13**, respectively. As expected, older patients (≥75 years) had higher cumulative incidences of both bladder-cancer and other-cause death compared with patients ages 66-74, regardless of cystoscopy frequency. Poor DS at diagnosis was also associated with higher cumulative incidences of both bladder-cancer and other-cause death compared with good DS, regardless of cystoscopy frequency. Among patients receiving ≥7 vs. <7 cystoscopies, the cumulative incidence of bladder-cancer death was the highest for patients with 1 or more comorbid conditions receiving ≥7 cystoscopies and the lowest for patients with no comorbid conditions receiving <7 cystoscopies. Compared with no comorbidities, having 1 or more comorbidities was also associated with higher cumulative incidence of other-cause death for both patients receiving ≥7 vs. <7 cystoscopies and ≥4 vs. <4 cystoscopies.

Figure 10. Cumulative incidence function for bladder-cancer and other-cause death



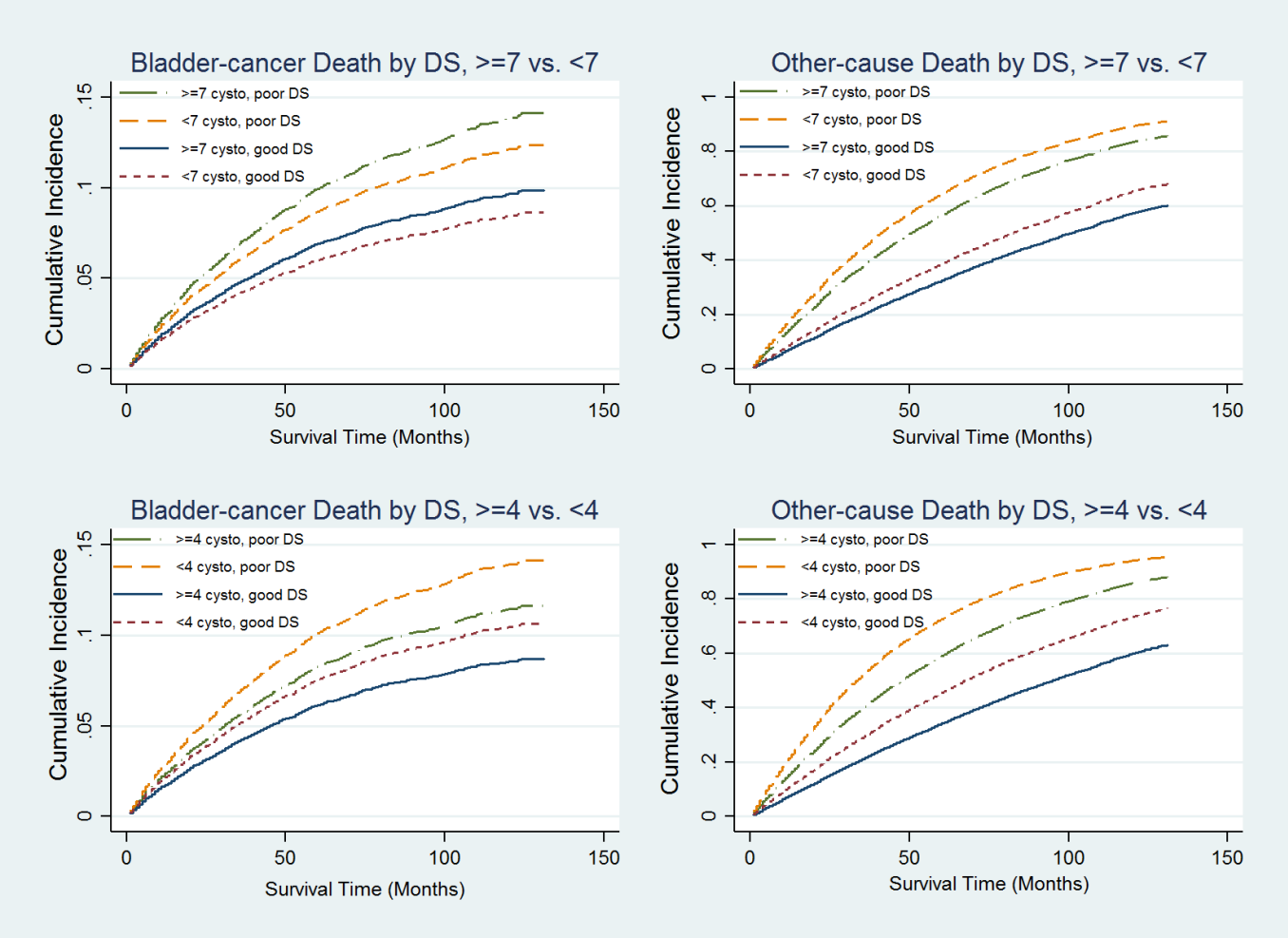
Note: Cumulative incidence functions (CIFs) generated from the propensity-score weighted Fine-Gray competing risk regression models to compare the cumulative incidence of bladder-cancer or other-cause death between patients who received ≥7 vs. <7 and ≥4 vs. <4 cystoscopies.

Figure 11. Cumulative incidence functions for bladder-cancer and other-cause death stratified by age (66-74 vs. 75+)



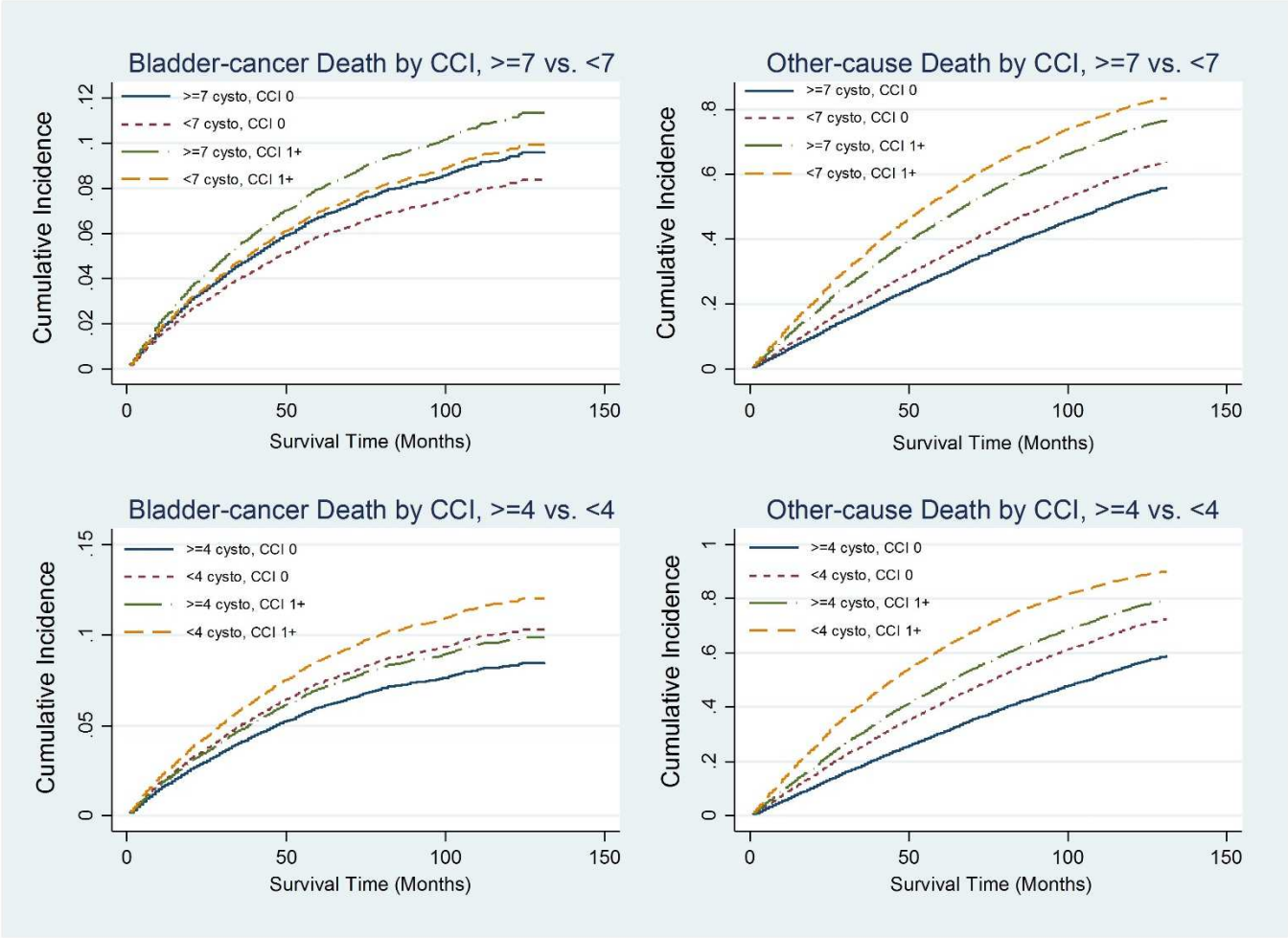
Note: Cumulative incidence functions (CIFs) generated from the propensity-score weighted Fine-Gray competing risk regression models to compare the cumulative incidence of bladder-cancer or other-cause death between patients who received ≥7 vs. <7 and ≥4 vs. <4 cystoscopies.

Figure 12. Cumulative incidence functions for bladder-cancer and other-cause death stratified by DS at diagnosis (poor vs. good DS)



Note: Cumulative incidence functions (CIFs) generated from the propensity-score weighted Fine-Gray competing risk regression models to compare the cumulative incidence of bladder-cancer or other-cause death between patients who received ≥7 vs. <7 and ≥4 vs. <4 cystoscopies.

Figure 13. Cumulative incidence functions for bladder-cancer and other-cause death stratified by comorbidities (0 vs. 1+)



Note: Cumulative incidence functions (CIFs) generated from the propensity-score weighted Fine-Gray competing risk regression models to compare the cumulative incidence of bladder-cancer or other-cause death between patients who received ≥7 vs. <7 and ≥4 vs. <4 cystoscopies.

Discussion

We found that receipt of both high-intensity (≥7 vs. <7 cystoscopies) and low-intensity surveillance (≥4 vs. <4 cystoscopies) were associated with a decrease in the hazard of all-cause death in the propensity-score adjusted survival models. Receipt of ≥7 cystoscopies was associated with an increase in the hazard of bladder-cancer death and a decrease in the hazard of other-cause death, which could be due to residual confounding by indication. Patients receiving ≥7 cystoscopies may have unobserved higher underlying risks of recurrence and/or progression, which are likely associated with both frequency of surveillance and cancer-specific survival and not accounted for by disease characteristics measured at baseline such as tumor stage and grade. Older patients (≥75 years) and those with poor DS at diagnosis also had higher cumulative incidences of both bladder-cancer and other-cause death compared with patients ages 66-74 and those with good DS, regardless of cystoscopy frequency.

Similar to previous studies, receipt of ≥4 cystoscopies was associated with a decrease in the hazard of bladder-cancer death (HR 0.81, 95%CI 0.72-0.91). A prior analysis of SEER-Medicare beneficiaries diagnosed with bladder cancer between 1992-2002 reported an even lower hazard of bladder-cancer death among patients receiving ≥4 vs. <4 cystoscopies (HR 0.61, 95%CI 0.47-0.79) using a similar propensity-score weighted Fine-Gray competing-risk regression model (Chamie et al. 2012). The study concluded that there was a statistically significant survival advantage of receiving ≥4 cystoscopies based on the findings for bladder-cancer mortality; however, it did not present any results for other-cause mortality. We found that receipt of ≥4 cystoscopies was significantly associated not only with lower bladder-cancer but also with lower other-cause mortality which has important implications for results interpretation. On the one hand, surveillance cystoscopy might be a proxy for improved access to care which in

turn resulted in better survival outcomes, including both lower bladder-cancer and other-cause mortality. On the other hand, it is also likely that the protective effect of receipt of ≥4 cystoscopies we found not only for bladder-cancer but also for other-cause mortality may be due to unmeasured healthy user bias, i.e., patients who are healthier based on unmeasured characteristics might be more likely to adhere to their recommended surveillance regimen and live longer.

Our study attempted to address confounding at baseline through propensity score (PS) weighting. We weighted patients receiving ≥7 vs. <7 and ≥4 vs. <4 cystoscopies on all observable and clinically meaningful characteristics at diagnosis, including age, comorbidity, disability status, tumor grade and stage, which were found to be strong predictors of receipt of ≥7 as well as ≥4 cystoscopies in Aim 1. After propensity-score weighting, baseline characteristics between patients receiving ≥7 vs. <7 and ≥4 vs. <4 cystoscopies were well-balanced, with SDM <10% for all covariates. By matching exposed and unexposed subjects based on observed covariates, PS methods can help reduce the bias in estimating treatment effects and the likelihood of confounding when analyzing nonrandomized, observational data (Haukoos and Lewis 2015). However, a major limitation of PS methods is that residual confounding may still exist if unmeasured or unobserved factors, not included in the PS model, influenced treatment selection (i.e., receipt of ≥7 or ≥4 cystoscopies).

We attempted to alleviate potential confounding by indication by including disability status at baseline which has not been addressed in previous studies (Chamie et al. 2012; Hollingsworth et al. 2010; Hollenbeck et al. 2009). In Aim 1 we found that patients with poor DS were significantly less likely to receive surveillance cystoscopy, even after controlling for other patient and disease characteristics, indicating that DS is an important factor influencing treatment

selection. However, it is likely that there are other unmeasured time-varying characteristics not included in the PS models, influencing both survival and surveillance use such as tumor recurrence and cancer progression, as well as progression of comorbid conditions contributing to other-cause mortality. We were not able to account for them, because the SEER registries do not conduct active follow-up of patients or collect information on recurrence, progression, or metastasis occurring after the initial diagnosis. Algorithms based on procedural codes to identify patients in SEER-Medicare with relapse or later metastatic disease have low sensitivity and are likely to miss a large percentage of patients with recurrences, particularly those who are older, and result in an incompletely and inaccurately classified cohort (Nordstrom et al. 2016; Earle et al. 2002; Warren et al. 2016; Warren and Yabroff 2015).

One potential limitation was the different follow-up periods for all-cause and cause-specific mortality. While data on time of death was available through December, 2016 for all-cause mortality, data on cause of death was available only until December, 2013; therefore, we used a shorter follow-up period to assess bladder-cancer and other-cause mortality. Similar to other studies (Chamie et al. 2012; Chamie et al. 2011), we restricted the study sample to patients who were alive and eligible for surveillance cystoscopy during the first two years after diagnosis to avoid immortal time bias (Suissa 2008). However, this exclusion may have resulted in potential selection bias and reduces generalizability to all NMIBC patients. We found that patients who underwent total cystectomy during the first two years after diagnosis were more likely to be younger, married, have secondary cancers, undifferentiated T1 tumor, no comorbidities, and good DS compared with patients who did not undergo total cystectomy. On the other hand, patients who died during the first two years after diagnosis were more likely to be

older, have history of Medicaid state buy-in in the 12 months prior to diagnosis, prior history of cancer, one or more comorbidity, and poor DS compared with those who did not die.

Even though we were not able to draw causal inference about the effect of surveillance cystoscopy use on survival due to potential unmeasured confounding, the findings from this observational study have several implications for future research. It is important that survival studies using competing-risk regression analysis report estimates for both disease-specific and other-cause mortality in order to assess whether the estimated effect on survival can be appropriately attributed to treatment or might be due to residual healthy user bias, even after applying methods such as PS weighting. Controlling for disability/functional status in observational studies of treatment effects can also help reduce potential confounding by indication.

Higher bladder-cancer mortality in older patients and those with poor disability status, regardless of frequency of cystoscopy, warrants further research. Randomized controlled trials as well as accounting for time-varying confounding are needed to assess the survival benefits and potential harms of different frequencies of surveillance among older patients with NMIBC, including impact on disease progression, survival, and health-related quality of life.

CHAPTER 6: COST-EFFECTIVENESS OF RISK-STRATIFIED SURVEILLANCE FOR OLDER PATIENTS WITH NON-MUSCLE-INVASIVE BLADDER CANCER IN THE UNITED STATES

Overview

Purpose: To compare the cost-effectiveness of different surveillance frequencies for non-muscle-invasive bladder cancer (NMIBC) based on risk-stratification or uniform cystoscopic evaluation for all patients. Additionally, we quantified trade-offs in terms of recurrent cases detected, progressed cases averted, deaths averted, and false positive cases averted under each surveillance strategy.

Methods: We developed a patient-level simulation model to compare three different US surveillance recommendations: a uniform approach of cystoscopy every three months for 2 years and less frequent thereafter, low-intensity and high-intensity risk-stratified approach (RSA) based on the 2016 American Urological Association guidelines. We projected downstream outcomes and costs from a healthcare sector and societal perspectives for a hypothetical cohort of 100,000 NMIBC patients aged ≥66 years. The time horizon of the model was five years with three-month cycles. Uncertainty in the input parameters was evaluated using probabilistic sensitivity analysis (PSA). All costs and utilities were discounted.

Results: The uniform recommendations had an incremental cost-effectiveness ratio (ICER) of $79,290 per QALY gained, compared with the high-intensity RSA. The ICER for the high-intensity RSA was $58,852 per quality-adjusted life years (QALY) gained, compared with the low-intensity RSA. The number of detected and progressed cases per person across the three

strategies ranged between 0.659-0.669 and 0.039-0.045, respectively, with more recurrences being undetected and progressing under the low-intensity RSA. False positives occurred in about 1 in 2 patients undergoing surveillance for five years under the uniform and the high-intensity RSA strategies, compared with 1 in 3 under the low-intensity RSA. In PSA the low-intensity RSA had a higher probability of being cost-effective compared with both the uniform and high-intensity RSA strategies under willingness-to-pay thresholds up to $120,000 per QALY gained.

Conclusion: Findings from our study suggest that intermediate-risk patients with NMIBC may benefit from less frequent surveillance than high-risk patients. The small differences in outcomes also highlight the importance of considering patient preferences for different trade-offs such as progressed cases vs. FP averted.

Key words: cost-effectiveness, non-muscle-invasive bladder cancer, surveillance, risk-stratification, simulation modelling

Introduction

Due to high recurrence rates, intensive surveillance strategies, and expensive therapies, the economic burden of bladder cancer is substantial (Svatek et al. 2014; Noyes, Singer, and Messing 2008). In the United States, the annual national cost of bladder cancer care in 2010 was estimated to be $3.98 billion and is expected to rise to $5 billion by 2020 (Mariotto et al. 2011). Surveillance and the management of recurrences accounted for approximately 60% of the lifetime cost of bladder cancer (Avritscher et al. 2006).

Most incident cases (70%) are non-muscle-invasive bladder cancer (NMIBC) at the time of presentation but these patients have highly variable risks of recurrence of NMIBC and progression to the potentially lethal phenotype of muscle-invasive (≥T2) disease (MIBC) (Burger

et al. 2013; Sylvester et al. 2006). Given the high recurrence rates, regular surveillance cystoscopy is recommended for patients with NMIBC of all ages to detect potential recurrences or progression to MIBC. Surveillance recommendations from existing clinical guidelines for NMIBC are largely consensus-based and vary across different professional societies and countries, with approaches ranging from intensive one-size-fits-all surveillance schedules, historically in the United States, to patient-level risk-stratification, historically in Europe (Chang et al. 2016; Hall et al. 2007; Power and Izawa 2016; Babjuk et al. 2016; Babjuk et al. 2013; Babjuk et al. 2011). In June 2016, the American Urological Association (AUA) and the Society of Urologic Oncology (SUO) released updated guidelines for diagnosis and treatment of NMIBC recommending a risk-stratified approach to surveillance cystoscopy, based on known risks for recurrence and progression (Chang et al. 2016). However, surprisingly, little is known about how different frequencies of surveillance affect patient outcomes and costs. Surveillance approaches following active treatment for cancer have been identified as the highest priority topic for cancer-related comparative effectiveness research (Greenberg et al. 2013). To date, no large randomized trials or observational studies have examined directly the trade-offs in terms of benefits, potential harms, and costs associated with different surveillance approaches.

Given the high financial and disease burden of bladder cancer as the US population ages, it is important to understand the implications of surveillance frequency and risk-stratification for providing high-quality, high-value care for older NMIBC patients. The purpose of this study was to compare the cost-effectiveness of different frequencies of surveillance cystoscopy for NMIBC under the 2016 AUA/SUO guidelines, using risk-stratification, and the historic US recommendations of uniform cystoscopic evaluation for all patients, regardless of their risks for recurrence and progression. Additionally, we quantified trade-offs in terms of recurrent cases

detected, progressed cases averted, deaths averted, and false positive cases averted under each surveillance strategy.

Methods

Model Overview

We developed a patient-level simulation model using TreeAge Pro 2017, R2.1 (TreeAge Software, Inc., Williamstown, MA, USA) comparing three different surveillance strategies: historic uniform US guidelines, low-intensity risk stratified approach (RSA), and high-intensity risk-stratified approach (RSA) based on the 2016 AUA/SUO guidelines (Table **12**) (Hall et al. 2007; Chang et al. 2016). The model had a time horizon of five years with three-month cycles, following the recommended total follow-up period in the guidelines (Chang et al. 2016). We projected downstream outcomes and costs for a hypothetical closed cohort of 100,000 Medicare patients with non-muscle invasive bladder cancer (NMIBC).

Table 12. Surveillance strategies compared in the cost-effectiveness model

	Historical US Guidelines	**Low-intensity RSA**			**High-intensity RSA**		
	All	**Low Risk**	**Intermediate Risk**	**High Risk**	**Low Risk**	**Intermediate Risk**	**High Risk**
Risk group definition	n/a	Low grade solitary Ta ≤3 cm	• Recurrence within 1 year, low grade Ta • Solitary low grade Ta >3 cm • Low grade Ta, multifocal • High grade Ta, ≤3 cm • Low grade T1	• High grade T1 • High grade Ta, >3 cm (or multifocal) • Any CIS • Any variant histology • Any LVI • Any high grade prostatic urethral involvement	Low grade solitary Ta ≤3 cm	• Recurrence within 1 year, low grade Ta • Solitary low grade Ta >3 cm • Low grade Ta, multifocal • High grade Ta, ≤3 cm • Low grade T1	• High grade T1 • High grade Ta, >3 cm (or multifocal) • Any CIS • Any variant histology • Any LVI • Any high grade prostatic urethral involvement
Cystoscopy frequency	Every 3 months for 2 years	3 months	3 months	Every 3 months for 2 years	3 months	Every 3 months for 2 years	Every 3 months for 2 years
	Every 6 months for 2 years	9 months	Every 6 months for 2 years	Every 6 months for years 3 and 4	6 months	Every 6 months for years 3 and 4	Every 6 months for years 3 and 4
	once/year	once/year	once/year	once/year	once/year	once/year	once/year
Total follow-up	5 years	5 years	5 years	5 years	5 years	5 years	5 years
Total cystoscopies	13	6	7	13	6	13	13

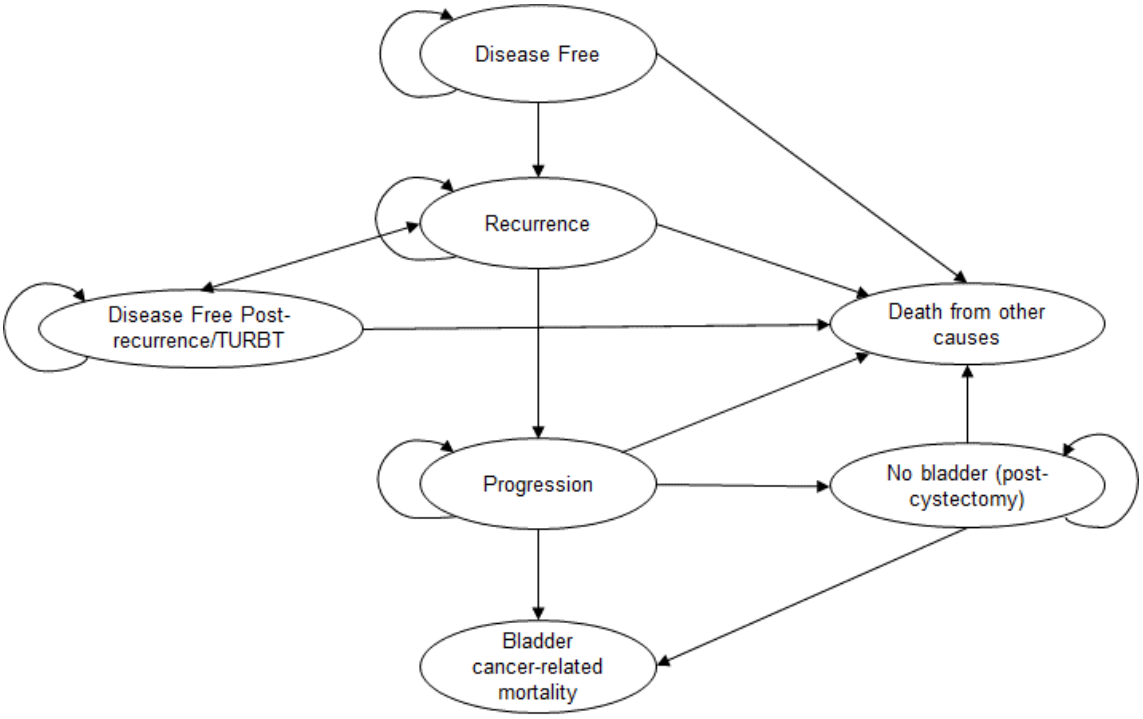
Abbreviation: RSA, Risk-stratified approach

Costs and outcomes (quality-adjusted life years) were discounted at an annual rate of 3%. Costs were evaluated from the healthcare sector and societal perspectives. All costs were inflated to 2017 USD, using the Consumer Price Index (CPI) for medical care (BLS 2018).

Figure **14** outlines the structure and flow of the model. Patients aged ≥66 years entered the "disease-free" state following an initial transurethral resection of the bladder tumor

(TURBT). At each 3-monthly model cycle the patient could experience a bladder cancer recurrence. If the recurrence was detected, the patient underwent an additional TURBT and returned to a disease-free state. However, if the recurrence was not detected, then the patient was at risk of progression and would undergo further treatment (cystectomy) once this progression was eventually detected. Death from background mortality could occur from any health state and bladder cancer mortality could occur only from the progression state or post-cystectomy state.

Figure 14. Disease state transition diagram.



Note: Patients enter the "disease-free" state following an initial transurethral resection of the bladder tumor (TURBT). At each 3-monthly model cycle the patient may experience a bladder cancer recurrence. If the recurrence is detected, the patient will undergo a further TURBT and return to a disease-free state. However, if the recurrence is not detected, then the patient will be at risk of progression and will have to undergo further treatment (cystectomy) once this progression is eventually detected. Death from background mortality can occur from any health state and bladder cancer mortality can occur only from the progression state or post-cystectomy state.

Model Assumptions

We made several assumptions based on our literature review:

- Only one recurrence could develop per surveillance cycle;
- Progression could occur only after recurrence (De Bekker-Grob et al. 2009; Van Kessel et al. 2013);
- Low-risk patients who experienced initial recurrence were re-classified as intermediate-risk (Chang et al. 2016), while patients who started as intermediate- or high-risk remained in the same risk group for the duration of the five-year model time horizon;
- If the tumor progressed to muscle invasive bladder cancer and was detected, patients received cystectomy (De Bekker-Grob et al. 2009; Van Kessel et al. 2013);
- After cystectomy, patients were no longer eligible for or underwent further surveillance (De Bekker-Grob et al. 2009; Van Kessel et al. 2013);
- Only patients with tumor progression at a certain moment or those in the post-cystectomy state were at higher risk for death than the background mortality rate (De Bekker-Grob et al. 2009; Van Kessel et al. 2013).

For surveillance stopped after 5 years (since 2016 AUA/SUO guideline recommendations), the decision whether to continue or stop routine follow-up cystoscopy after five years should be based on shared-decision making between the patient and clinician (Chang et al. 2016).

Surveillance Strategies

We modelled three surveillance strategies based on the 2016 AUA/SUO risk-stratified guidelines (Chang et al. 2016) and the historic uniform US guidelines (Hall et al. 2007) (**Table 12**):

1) Low-intensity RSA:
    a. Low-risk group: surveillance cystoscopy at three months, nine months, and then annually thereafter for five years.
    b. Intermediate-risk group: surveillance cystoscopy at three months, every six months for two years, and annually thereafter;
    c. High-risk group: surveillance cystoscopy every three months for two years, every six months for years 3 and 4, and annually thereafter;
2) High-intensity RSA:
    a. Low-risk group: surveillance cystoscopy at three months, six months and annually thereafter for five years;
    b. Intermediate-risk group: surveillance cystoscopy every three months for two years, every six months for years 3 and 4 and annually thereafter;
    c. High-risk group: surveillance cystoscopy every three months for two years, every six months for years 3 and 4 and annually thereafter;
3) Historic uniform approach: cystoscopy every 3 months for 2 years, every 6 months for 2 years and annually thereafter.

Patient Population and Risk Groups

We simulated 100,000 Medicare patients with NMIBC aged ≥66 years and stratified them in three risk groups for recurrence and progression, following the 2016 AUA/SUO surveillance guidelines (Table **12**). Low-risk patients included those with low grade solitary Ta tumor with a diameter ≤ 3cm. Since SEER does not collect data on papillary urothelial neoplasms of low malignant potential (PUNLMP), they were not included as part of the low-risk classification in the model, whereas they are included in the 2016 AUA/SUO Guideline low-risk

group (SEER 2016). Patients classified as low-risk at baseline who experienced cancer recurrence were re-classified as intermediate risk for subsequent cycles of the model. The intermediate-risk group also included patients with any of the following characteristics at baseline: solitary low-grade Ta tumor with a diameter >3cm; low grade Ta multifocal tumor; high-grade Ta tumor with a diameter ≤ 3cm; or low grade T1 tumor. The high-risk group included patients who had at baseline any of the following: high grade T1; high grade Ta, >3 cm (or multifocal); any Tis; any variant histology; any lymphovascular invasion; or any high grade prostatic urethral involvement.

Probabilities of the baseline characteristics used to stratify patients into three risk groups were obtained from the SEER Patient Entitlement and Diagnosis Summary File (PEDSF). Probabilities of tumor size and presence of solitary vs. multifocal tumors were obtained from the European Organization for Research and Treatment of Cancer (EORTC) trials (Sylvester et al. 2006). Probabilities of remaining patient characteristics used in the risk-stratification algorithm (i.e., any lymphovascular invasion; any variant histology; any high grade prostatic urethral involvement) were informed by review of the published literature (Table **13**).

Transition Probabilities, Utility, and Cost Data

Model parameters, values, and distributions are presented in Table 13. Annual probabilities of recurrence and progression over five years for the three risk groups were obtained from the EORTC risk tables (Sylvester et al. 2006). Bladder cancer and other cause mortality data were obtained from the published literature and CDC life tables, respectively. Test characteristics of cystoscopy (sensitivity and specificity to detect bladder cancer) were obtained from a large systematic review and meta-analysis of published studies (Blick et al. 2012). All

annual probabilities of death, recurrence, and progression were converted to three-month state transition probabilities (Fleurence and Hollenbeak 2007).

Utilities associated with a given disease state, risk group, or procedure were obtained from the published literature. Costs of medical procedures were obtained from the Medicare claims data and review of the published literature (Table **13**). Societal costs also included transportation costs and costs due to productivity loss from medical procedures (Sanders et al. 2016). Transportation costs were calculated based on the distance traveled to the treating facility, estimated in Aim 1, and average national taxi fares per mile, since patients might not be able to drive for at least 24 hours after their procedure. Costs due to productivity loss were calculated based on the average length of stay associated with cystoscopy, cystectomy, and TURBT visits, and the national average wage per hour in the United States (BLS 2017).

Table 13. Model input parameters

VARIABLES	BASE-CASE VALUE [a]		DISTRIBUTION	SOURCE
Patient characteristics				
Sex (males)	0.75		β	SEER-Medicare (PEDSF) – Aim 1
Age	Male	Female		SEER-Medicare (PEDSF) – Aim 1
66-74 years	0.36	0.33	Dirichlet	
75-84 years	0.47	0.45	Dirichlet	
≥85 years	0.17	0.22	Dirichlet	
Stage at diagnosis (T Category)				SEER-Medicare (PEDSF) – Aim 1
Ta	0.62		Dirichlet	
Tis	0.08		Dirichlet	
T1	0.30		Dirichlet	
Tumor grade				SEER-Medicare (PEDSF) – Aim 1
G1: Well differentiated (low grade)	0.19		Dirichlet	
G2: Moderately differentiated (intermediate grade)	0.39		Dirichlet	
G3: Poorly differentiated (high grade)	0.21		Dirichlet	

G4: Undifferentiated (high grade)			0.21			Dirichlet	
Tumor size (cm)							EORTC trial data (Sylvester et al. 2006)
≤3 cm			81.8%			β	
>3 cm			18.2%			β	
Solitary tumor			57%			β	EORTC trial data (Sylvester et al. 2006)
Any variant histology among T1			10%			β	(Black, Brown, and Dinney 2009)
Any lymphovascular invasion (LVI) among T1			13.8%			β	(Lotan et al. 2005; Kikuchi et al. 2009)
Any high grade prostatic urethral involvement			10%			β	(Palou et al. 2012)
Three-month state transition probabilities							
3-month probability of recurrence	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Dirichlet	EORTC trial data (Sylvester et al. 2006)
Low-risk	3.98%;	2.90%;	2.37%;	2.03%;	1.84%		
Intermediate-risk	6.63%;	5.06%;	4.17%;	3.45%;	3.03%		
High-risk	11.26%;	8.53%;	6.61%;	5.42%;	4.72%		
3-month probability of progression given recurrence	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Dirichlet	EORTC trial data (Sylvester et al. 2006)
Low-risk	1.26%;	0.86%;	2.82%;	2.47%;	2.18%		
Intermediate-risk	3.78%;	7.51%;	8.15%;	9.27%;	10.18%		
High-risk	21.70%;	25.77%;	26.84%;	33.69%;	40.39%		
Conditional probability of death after undetected progression to MIBC							(De Bekker-Grob et al. 2009)
3-month probability of death during year 1			12.0%			Dirichlet	
3-month probability of death during year 2			2.5%			Dirichlet	
3-month probability of death during year 4			2.1%			Dirichlet	
Bladder-cancer death following cystectomy among high-risk patients							(Lee et al. 2007)
3-month probability of death			2%			β	
Other-cause mortality by age group						Dirichlet	CDC life tables 2014 (Arias et al. 2017)
3-month probability of death (66-74 years)			3.3%				
3-month probability of death (75-84 years)			9.0%				
3-month probability of death (≥85 years)			11.5%				
Test characteristics							
Flexible cystoscopy: sensitivity			0.98 (95% CI: 0.94 – 0.99)			β	(Blick et al. 2012)
Flexible cystoscopy: specificity			0.94 (95% CI: 0.92 – 0.96)			β	(Blick et al. 2012)

Costs (2017 USD)				
Healthcare Sector	Medicare Payment	Out-of-pocket Costs		
Cystoscopy (HCPCS: 52000)	$365 (SD, $566)	$84 (SD, $147)	γ	Medicare claims, BLS CPI (2017)
Cystectomy (HCPCS: 51550–51597)	$14,669 (SD, $20,268)	$188 (SD, $223)	γ	Medicare claims, BLS CPI (2017)
TURBT (HCPCS: 52234, 52235, 52240)	$1,347 (SD, $2,015)	$273 (SD, $305)	γ	Medicare claims, BLS CPI (2017)
Societal				
Transportation costs	$45 (SD, $20)		γ	Medicare claims, ("Taxi Fares in the US" 2017)
Productivity loss due to cystoscopy	$24 (SD, $10; 1 hour x $24)		γ	(BLS 2017; Mayo Clinic 2018)
Productivity loss due to TURBT	$1,152 (SD, $100; 48 hours x $24)		γ	(BLS 2017; Mayo Clinic 2018)
Productivity loss due to cystectomy	$3,456 (SD, $200; 144 hours x $24)		γ	(BLS 2017; Mayo Clinic 2018)
Utility or Disutility				
Surveillance cystoscopy	-0.003		Uniform (±10%)	(Kulkarni et al. 2007; Kulkarni et al. 2009)
TURBT	-0.10		Uniform (±10%)	(Kulkarni et al. 2007; Kulkarni et al. 2009)
Cystectomy	-0.20		Uniform (±10%)	(Kulkarni et al. 2007; Kulkarni et al. 2009)
Post-cystectomy	-0.04		Uniform (±10%)	(Kulkarni et al. 2007; Kulkarni et al. 2009)
Low-risk NMIBC	0.98		β	(Zhang, Denton, and Nielsen 2013)
Intermediate-risk NMIBC	0.95		β	(Zhang, Denton, and Nielsen 2013)
High-risk NMIBC	0.93		β	(Zhang, Denton, and Nielsen 2013)
Muscle-invasive bladder cancer (progression)	0.80		β	(Zhang, Denton, and Nielsen 2013)

Model Outcomes

Model outcomes per patient and for the entire cohort were assessed over five years of surveillance under each strategy and included total quality-adjusted life years (QALYs), total cancer recurrences detected, total progressed cases detected, total deaths, total false positive (FP) cases, and total discounted costs from the healthcare sector perspective, as the primary analysis, and from the societal perspective, as a secondary analysis.

Base-case Cost-effectiveness Analysis

To evaluate costs per outcome achieved, we ranked the three strategies from least to most expensive and compared them sequentially. We calculated incremental differences in total costs, QALYs, total deaths, total detected recurrent cases, total detected progressed cases, and total FP cases. We then calculated incremental cost-effectiveness ratios (ICERs) for each strategy as the additional cost divided by the change in outcome (QALY gained, death averted recurrence detected, progression averted, FP averted) compared with the next less expensive alternative, removing any dominated strategies from the next sequential comparison (Drummond, Stoddard, and Torrance 2005). Costs and outcomes were discounted at an annual rate of 3%. We used the commonly accepted in the United States threshold of $50,000 per QALY gained and a higher threshold of $100,000 per QALY gained to compare the cost-effectiveness of the three surveillance strategies (Neumann, Cohen, and Weinstein 2014).

Sensitivity Analyses

We performed probabilistic sensitivity analysis with 1,000 Monte Carlo simulations over a range of plausible values of the ICERs per QALY gained, given uncertainty in the input parameters, and graphed these simulations as ICER planes, which represents the effectiveness difference per patient on the horizontal axis plotted against the difference in costs per patient on the vertical axis (Briggs, Sculpher, and Claxton 2006). We used beta distribution for binomial data, including utilities and test characteristics (sensitivity and specificity), Dirichlet distribution for multinomial data, and gamma distribution for costs (Briggs, Sculpher, and Claxton 2006). Input parameters for which ranges or 95% CI were not reported in the literature were varied by $\pm$10%, using a uniform distribution (Table **13**).

We also evaluated uncertainty using cost-effectiveness acceptability curves (CEACs). The CEACs allow decision makers to compare the probability of a strategy being cost-effective under a range of different willingness-to-pay (WTP) thresholds. We constructed CEACs by plotting the probability that the estimated cost-effectiveness ratio (additional cost per QALY gained) for each surveillance strategy falls below specified values of WTP using the net-benefit framework (Briggs, Sculpher, and Claxton 2006). The net-benefit framework relies on the net monetary benefit or net health benefit statistics to overcome some of the problems associated with ICERs such as having ICERs of the same sign in opposite quadrants of the cost-effectiveness plane (Briggs, Sculpher, and Claxton 2006).

Results

We found that, among the simulated cohort of 100,000 Medicare patients with NMIBC, 17% were at low risk, 43% were at intermediate risk, and 40% were at high risk for disease recurrence and progression. When comparing the health outcomes per patient over 5 years of surveillance under each strategy (Table **14**, Table **15**), the historic uniform recommendations detected 0.669 recurrent cases per patient—0.039 more than the high-intensity RSA and 0.010 more than the low-intensity RSA. The uniform recommendations and the high-intensity RSA detected the same number of progressed cases per patient (0.039), while the low-intensity RSA detected a slightly higher number of progressed cases per patient—0.045, because more recurrences went undetected and progressed to muscle-invasive disease under this strategy. However, potential harms associated with frequent surveillance were the resulting false positive cases from cystoscopy evaluation. Under the uniform recommendations, 0.478 FP cases occurred per patient, followed by 0.437 under the high-intensity RSA, and 0.372 under the low-intensity RSA. In other words, about 1 in 2 people undergoing surveillance for five years under the

uniform and the high-intensity RSA recommendations will have a FP, compared with 1 in 3 people undergoing surveillance under the low-intensity RSA. Considering the high average age at diagnosis (77 years) of the simulated cohort, most deaths occurred due to other-cause mortality. The number of deaths was high across all three strategies, ranging from 0.639 to 0.643 deaths per patient.

Table 14. Base-case Results per Patient Undergoing Surveillance under Each Strategy over 5 Years. Results in parentheses represent value for the entire simulated cohort (N=100,000).

Strategy	Total costs*	Total QALYs	Total deaths	Δ costs*	Δ QALYs	Δ deaths	ICER/ QALY gained	ICER/ death averted
Healthcare sector perspective								
Low-intensity Risk-stratified	$3,675 ($367,466,765)	2.86 (285,902)	0.643 (64,335)	referent	referent	referent	referent	referent
High-intensity Risk-stratified	$4,370 ($437,021,263)	2.87 (287,084)	0.640 (63,987)	$696 ($69,554,498)	0.012 (1,182)	-0.003 (-348)	$58,852	$200,077
Historic Uniform	$4,633 ($463,334,755)	2.87 (287,415)	0.639 (63,895)	$263 ($26,313,491)	0.003 (332)	-0.001 (-92)	$79,290	$284,948
Societal perspective								
Low-intensity Risk-stratified	$4,687 ($468,677,708)	2.86 (285,902)	0.643 (64,335)	referent	referent	referent	referent	referent
High-intensity Risk-stratified	$5,466 ($546,562,546)	2.87 (287,084)	0.640 (63,987)	$779 ($77,884,838)	0.012 (1,182)	-0.003 (-348)	$65,900	$224,039
Historic Uniform	$5,747 ($574,697,939)	2.87 (287,415)	0.639 (63,895)	$281 ($28,135,393)	0.003 (332)	-0.001 (-92)	$84,780	$304,677

Note: QALYs – quality-adjusted life years; ICER – incremental cost-effectiveness ratio; Δ – change.
*All costs were inflated to 2017 USD. Costs and QALYs were discounted at 3% annual discount rate.

Table 15. Base-case Results per Patient Undergoing Surveillance under Each Strategy over 5 Years. Results in parentheses represent value for the entire simulated cohort (N=100,000) (cont.)

Strategy	Total costs*	Total detected recurrent cases	Total detected progressed cases	Total FPs	Δ costs*	Δ recurrent cases	Δ progressed cases	Δ FP cases	ICER/ recurrence detected	ICER/ progression averted	ICER/ FP averted
Healthcare sector perspective											
Low-intensity Risk-stratified	$3,675 ($367,466,765)	0.659 (65,861)	0.045 (4,544)	0.372 (32,705)	referent	referent	referent	referent	referent	referent	referent
High-intensity Risk-stratified	$4,370 ($437,021,263)	0.669 (66,852)	0.039 (3,914)	0.437 (43,705)	$696 ($69,554,498)	0.010 (991)	-0.006 (-631)	0.110 (11,000)	$70,200	$110,278	dominated†
Historic Uniform	$4,633 ($463,334,755)	0.669 (66,892)	0.039 (3,916)	0.478 (47,760)	$263 ($26,313,491)	0.000 (40)	0.000 (2)	0.041 (4,055)	$652,180	dominated†	dominated†
Societal perspective											
Low-intensity Risk-stratified	$4,687 ($468,677,708)	0.659 (65,861)	0.045 (4,544)	0.372 (32,705)	referent	referent	referent	referent	referent	referent	referent
High-intensity Risk-stratified	$5,466 ($546,562,546)	0.669 (66,852)	0.039 (3,914)	0.437 (43,705)	$779 ($77,884,838)	0.010 (991)	-0.006 (-631)	0.110 (11,000)	$78,607	$123,486	dominated†
Historic Uniform	$5,747 ($574,697,939)	0.669 (66,892)	0.039 (3,916)	0.478 (47,760)	$281 ($28,135,393)	0.000 (40)	0.000 (2)	0.041 (4,055)	$697,335	dominated†	dominated†

Note: QALYs – quality-adjusted life years; ICER – incremental cost-effectiveness ratio; FP – false positive; Δ – change.
*All costs were inflated to 2017 USD. Costs and QALYs were discounted at 3% annual discount rate.
† Higher cost, lower effectiveness (e.g., fewer progressed cases averted, fewer FP averted).

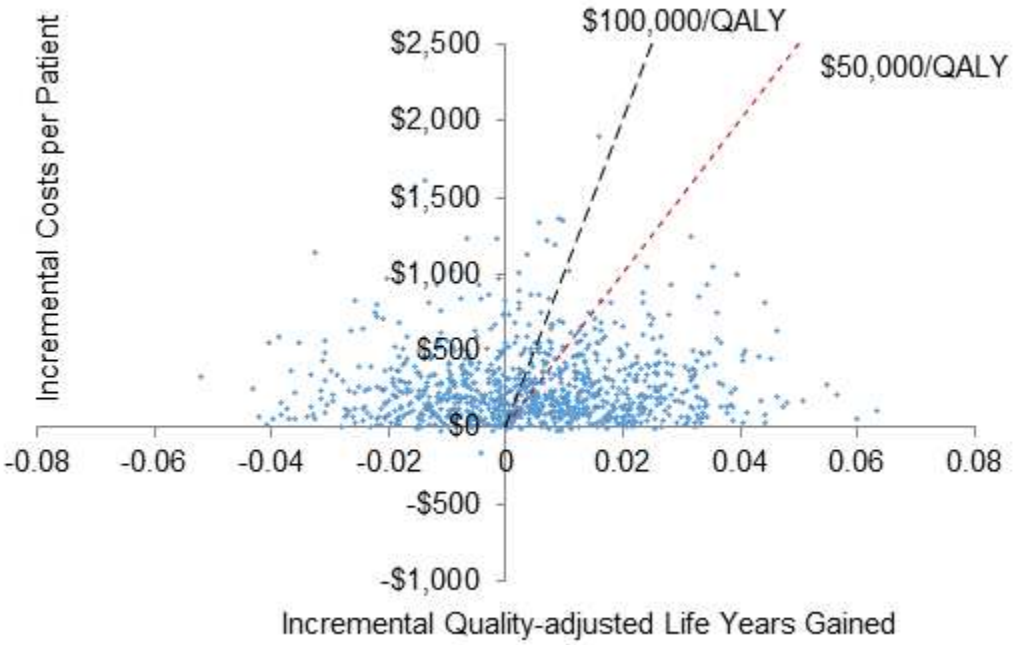
In the primary analysis from the healthcare sector perspective, the uniform recommendation was the highest cost strategy ($4,633 per patient), followed by the high-intensity RSA ($4,370 per patient), and low-intensity RSA ($3,675 per patient). In the base-case cost-effectiveness analysis (Table **14**, Table **15**), we found that after five years of surveillance, compared with the high-intensity RSA, the uniform recommendations were associated with an additional cost of $263 per patient and incremental benefits of 0.003 more QALYs gained per person and 0.001 deaths averted per person. The incremental differences in QALYs gained and deaths averted per person between the two strategies were minimal because the majority of patients (83%) were intermediate or high risk and received 13 cystoscopies over five years under both the high-intensity RSA and the uniform approach. This resulted in ICERs for the historic uniform guidelines of $79,290 per QALY gained and $284,948 per death averted compared with the high-intensity RSA. Compared with the low-intensity RSA, the high-intensity RSA was associated with an additional cost of $696 per person and incremental benefits of 0.012 more QALYs gained per person and 0.003 deaths averted per person. The ICERs for the high-intensity RSA were $58,852 per QALY gained and $200,077 per death averted, compared with the low-intensity RSA. Both the uniform guidelines and the high-intensity RSA would be considered cost-effective compared with the next less expensive strategy under the higher threshold of $100,000 per QALY gained but not under the standard threshold of $50,000 per QALY gained. Similar results were reported using the societal perspective for costs, with total costs per person being ~$1,000 higher for each strategy than under the healthcare sector perspective (Table **14**).

Sensitivity Analysis

In the probabilistic sensitivity analysis using 1,000 Monte Carlo simulations (Figure **15** and Figure **16**.), after 5 years of surveillance, the historic uniform AUA recommendations were

more cost-effective than the high-intensity RSA in fewer than half of the simulations under the standard US threshold of $50,000 per QALY gained and in 54% of the simulations under a higher threshold of $100,000 per QALY gained. In 1% of the simulations the uniform recommendations were cost-saving, i.e., had a lower incremental cost and higher QALYs gained, compared with the high-intensity RSA (Figure **15**). The high-intensity RSA was more cost-effective than the low-intensity RSA in 47% of the simulations under the standard US threshold of $50,000 per QALY gained and in 55% of the simulations under a higher threshold of $100,000 per QALY gained, with 2% of the simulations being cost-saving (Figure **16**).

Figure 15. Incremental cost-effectiveness plane for the Historic AUA strategy (blue) compared with the High-intensity Risk-Stratified AUA strategy over 5 years of surveillance.



Note: Results in the upper right quadrant are cost-effective if they fall below the standard US threshold of $50,000/QALY gained. Results in the lower right quadrant indicate dominant cost-saving strategy (better outcomes at a lower cost). Approximately 47% of the 1,000 replications were cost-effective under the $50,000/QALY gained threshold and 54% were cost-effective under the $100,000/QALY gained threshold. 1% of the simulations were cost-saving.

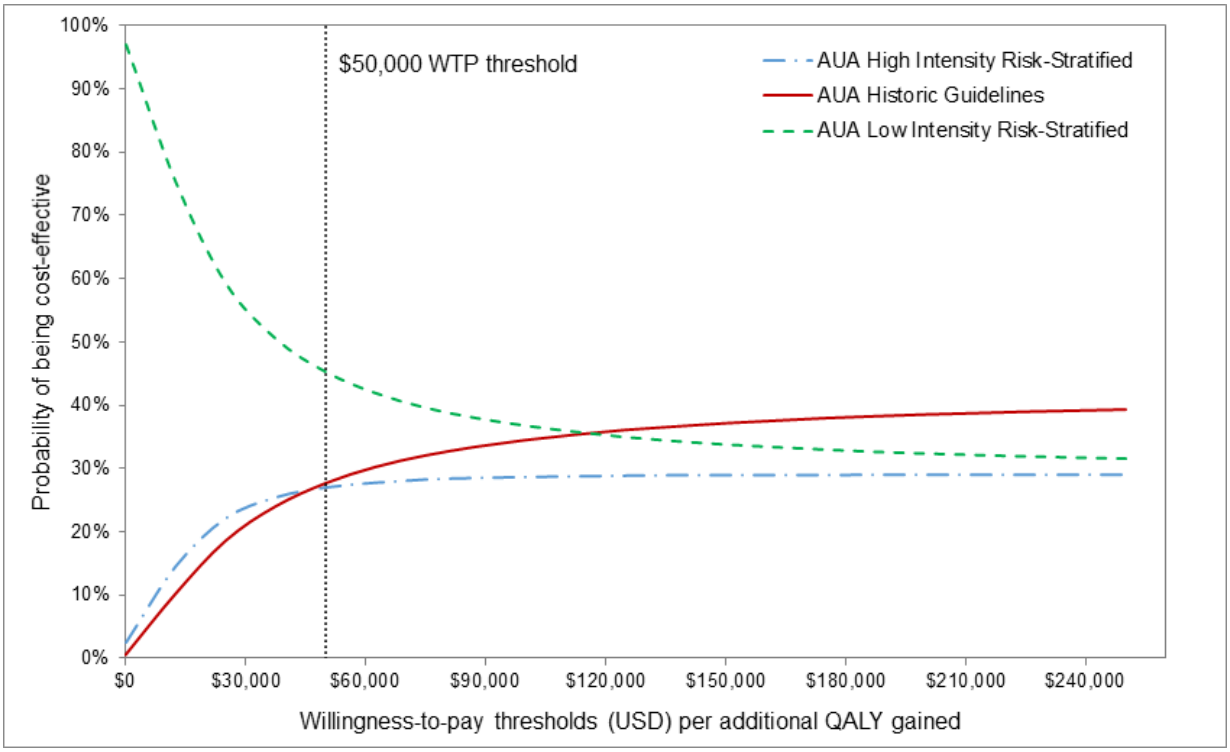
Figure 16. Incremental cost-effectiveness plane for the High-intensity Risk-Stratified AUA strategy (green) compared with the Low-intensity Risk-Stratified AUA strategy over 5 years of surveillance.

Note: Results in the upper right quadrant are cost-effective if they fall below the standard US threshold of $50,000/QALY gained. Results in the lower right quadrant indicate dominant cost-saving strategy (better outcomes at a lower cost). Approximately 47% of the 1,000 replications were cost-effective under the $50,000/QALY gained threshold and 55% were cost-effective under the $100,000/QALY gained threshold. 2% of the simulations were cost-saving.

We also evaluated uncertainty using CEACs which show the probabilities that each strategy would be considered cost-effective under various WTP per additional QALY gained thresholds (Figure **17**). The low-intensity RSA had a higher probability of being cost-effective compared with both the uniform recommendations and high-intensity RSA under WTP thresholds up to $120,000 per QALY gained. At the commonly accepted US threshold of $50,000 per QALY gained, the probability of being cost-effective was 45% for the low-intensity RSA, 28% for the uniform, and 27% for the high-intensity RSA recommendations. If decision makers were willing to pay more than $120,000 per QALY gained, then the historic uniform recommendations had the highest probability of being cost-effective, compared with the low-

intensity RSA or the high-intensity RSA. The high-intensity RSA had a lower probability of being cost-effective compared with the low-intensity RSA across the entire range of modeled WTP thresholds.

Figure 17. Cost-effectiveness acceptability curves



Note: Cost-effectiveness acceptability curves (CEACs) of 5-year surveillance for Medicare patients with non-muscle invasive bladder cancer (NMIBC) under each of the different recommendations from a US healthcare sector perspective. Each CEAC represents the probability that a strategy is cost-effective under different willingness-to-pay (WTP) thresholds for an additional quality-adjusted life year (QALY) gained. In the United States a cost-effectiveness threshold of $50,000 per QALY gained is commonly accepted. The AUA low-intensity risk-stratified approach had a higher probability of being cost-effective compared with the historic or high-intensity risk-stratified AUA recommendations under WTP thresholds up to $120,000 per QALY gained.

Discussion

This is the first cost-effectiveness study comparing different frequencies of surveillance cystoscopy for NMIBC under the 2016 AUA/SUO guidelines using risk-stratification and the historic US surveillance approach, recommending uniform cystoscopic evaluation for all

patients. We found that a sizeable proportion of patients (43%) were intermediate-risk, for which the current AUA/SUO guidelines do not have definitive recommendations on what the optimal frequency of surveillance should be (Chang et al. 2016). The historic uniform recommendations, high-intensity and low-intensity RSA were associated with different trade-offs in terms of QALYs gained, recurrent cases detected, progressed cases averted, deaths averted, false positive cases averted, and costs. The high-intensity RSA and the uniform approach had similar health outcomes and costs per patient because the majority of patients in the simulated cohort were at intermediate or high risk and received 13 cystoscopies over five years, as many as under the uniform approach. Compared with the high-intensity RSA, the low-intensity RSA was associated with a slightly higher number of deaths (348 per 100,000 people) and lower number of recurrences detected (991 per 100,000 people), which resulted in more undetected recurrences progressing to MIBC (631 per 100,000 people). However, for every progressed case, 17 FP cases were avoided under the low-intensity RSA, compared with the high-intensity RSA.

Patients with clinically similar disease may have heterogeneous preferences for the different trade-offs associated with the three surveillance strategies, particularly between progressed cases and FPs averted. While some patients might have a strong preference for detecting recurrences as early as possible and avoiding progression at all cost, therefore willingly undergo more frequent cystoscopic evaluation, others might prefer to minimize potential FP results, the physical discomfort, and costs associated with frequent surveillance. High false positive findings can result not only in unnecessary workups and costs but also in increased patient morbidity and anxiety (Shariat et al. 2008). Surveillance cystoscopy is also associated with patient anxiety and discomfort (Koo et al. 2017).

In probabilistic sensitivity analysis, the low-intensity RSA was more cost-effective than the high-intensity RSA and the uniform approach under a wide range of WTP thresholds up to $120,000 per QALY gained. If decision makers are willing to pay more than $120,000 per QALY gained, the historic uniform approach would be the more cost-effective one. Considering the high financial burden of bladder cancer, with surveillance and management of recurrences being the main drivers of the lifetime cost (Avritscher et al. 2006), a less intensive risk-stratified surveillance, particularly for patients at low- and intermediate-risk, may help decrease the costs of managing NMIBC.

The National Institute for Health and Care Excellence (NICE) in the United Kingdom recommends even less frequent surveillance schedule for low- and intermediate-risk NMIBC patients than the 2016 AUA/SUO guidelines. According to the NICE guidelines, low-risk patients should undergo cystoscopic follow-up at three and 12 months after diagnosis and be discharge to primary care within 12 months if they did not have a recurrence (National Collaborating Centre for Cancer 2015). For intermediate-risk patients cystoscopy is recommended at three months, nine months, 18 months, and annually thereafter for five years, for a total of six cystoscopic evaluations over five years, compared with 7-13 evaluations under the AUA/SUO risk-stratified approach (National Collaborating Centre for Cancer 2015; Chang et al. 2016).

Our study has several potential limitations. We did not consider PUNLMP as part of the low-risk classification algorithm because data on PUNLMP are not collected by the SEER registries (SEER 2016). As a result, the proportion of patients classified in our model as low-risk at baseline might be an underestimate of the low-risk group in a broader real-world population. The frequency of PUNLMP ranged from 14%-22% in hospital- or clinic-based studies to 36%-

39% in population-based studies (Schned et al. 2008). Patients who started as low-risk in our model were re-classified as intermediate-risk after initial recurrence, however, the risk-groups were not dynamically re-assessed and re-classified after subsequent recurrences. As a result, we might have also underestimated the proportion of high-risk patients over time, particularly among intermediate-risk patients with high grade Ta tumor who experienced a recurrence and patients with high grade tumor who failed Bacillus Calmette-Guérin (BCG) immunotherapy. BCG failures occurred in 35% of high grade cases, or in less than 13% of all NMIBC patients (Witjes 2006). We assumed that everyone who progressed to MIBC underwent cystectomy, whereas among Medicare patients there is variation in the strategies used to treat MIBC, including chemotherapy and radiotherapy (Schrag et al. 2005). Lastly, we used both the healthcare and societal perspectives to characterize medical, transportation costs, and productivity loss due to cystoscopy visits and management of recurrences and progression. However, our model did not follow the treatment trajectory beyond cystectomy of patients progressing to MIBC, which potentially includes combinations of different chemotherapy and immunotherapy regimens, particularly for patients with advanced disease (Flaig et al. 2018). As a result, the costs associated with delayed detection of progression and treatment of advanced disease may be higher. However, FPs may also be associated with downstream costs of additional workup, thus contributing to the cost of frequent surveillance.

Conclusion

Using a patient-level simulation model of a hypothetical cohort of 100,000 Medicare patients with NMIBC, we found that low-intensity risk-stratified surveillance over five years, with cystoscopy frequency increasing progressively with risk, was associated with small differences in QALYs, deaths, detected recurred and progressed cases, lower costs, and

substantially fewer false positives, compared with a more frequent high-intensity risk-stratified approach and uniform surveillance. Findings from our study suggest that intermediate-risk patients with NMIBC may benefit from less frequent surveillance than high-risk patients. The small differences in outcomes also highlight the importance of considering patient preferences for different trade-offs such as progressed cases vs. FP averted.

CHAPTER 7: STUDY LIMITATIONS, POLICY IMPLICATIONS, AND FUTURE DIRECTIONS

Summary of Findings

This dissertation examined the value of surveillance for older patients with NMIBC by assessing disease characteristics, surveillance patterns, and health-economic outcomes in a population-based cohort of Medicare patients with NMIBC. This is the first study to directly compare the cost-effectiveness of the 2016 AUA/SUO guidelines, which adopted risk-stratified surveillance, with historical US guidelines in which recommended uniform surveillance for all patients. Moreover, we examined patients' disability status at diagnosis as an important predictor of surveillance use and survival in NMIBC patients, which has not been previously investigated.

In Chapter 4, we found that NMIBC patients aged ≥85 years and those with poor DS and ≥3 comorbidities at diagnosis were least likely to undergo recommended (≥7 cystoscopies) and low-intensity (≥4 cystoscopies) surveillance over the first two years after diagnosis. As the age at diagnosis and the number of comorbid conditions increased, the odds of receiving recommended cystoscopy frequency, as well as the rate of cystoscopy, decreased. Findings from our study also indicated potential racial disparities in receipt of surveillance: after adjusting for clinical, demographic, and socioeconomic factors, non-Hispanic black patients were significantly less likely to receive cystoscopy across all three surveillance outcomes, compared with non-Hispanic white patients. At the same time, a higher proportion of patients with poor DS were black compared with those with good DS, suggesting greater functional decline among black patients.

In Chapter 5, we found that receipt of either high-intensity or low-intensity surveillance was associated with a decrease in the hazard of all-cause death in the propensity-score weighted survival models. Receipt of ≥7 cystoscopies was associated with an increase in the hazard of bladder-cancer death and a decrease in the hazard of other-cause death, which could be due to potential unmeasured confounding. Older patients (≥75 vs. 66-74 years) and those with poor disability status at diagnosis had higher cumulative incidence of both bladder-cancer and other-cause death, regardless of frequency of cystoscopy. Our study attempted to address endogenous selection into cystoscopy by healthier patients through propensity score (PS) weighting and accounting for disability status at baseline which has not been addressed in previous studies. We balanced patients receiving ≥7 vs. <7 and ≥4 vs. <4 cystoscopies on all observable and clinically meaningful characteristics at diagnosis, including age, comorbidity, DS, tumor grade and stage, which were found to be strong predictors of receipt of ≥7 as well as ≥4 cystoscopies. However, it is likely that there are other unmeasured time-varying characteristics not included in the PS models, influencing both survival and surveillance use such as tumor recurrence, response to treatment, and cancer progression.

In Chapter 6, using a patient-level simulation model of a hypothetical cohort of 100,000 Medicare patients with NMIBC, we found that low-intensity risk-stratified surveillance over five years, with cystoscopy frequency increasing progressively with risk, was associated with a slightly higher number of deaths and lower number of recurrences detected, which resulted in more undetected recurrences progressing to MIBC but lower costs and substantially fewer false positives, compared with a more frequent high-intensity risk-stratified approach and uniform surveillance.

This dissertation makes several important contributions to the literature. First, no previous studies have accounted for DS at diagnosis as a predictor of surveillance use, nor has DS been incorporated in survival models examining surveillance (Chamie et al. 2012; Hollingsworth et al. 2010; Hollenbeck et al. 2009). Furthermore, no studies have examined recent trends in NMIBC surveillance in a broader cohort of NMIBC patients with both high and low-grade tumors. Lastly, this is the first study to investigate the cost-effectiveness of the 2016 AUA/SUO guidelines using risk-stratification, compared with the historic US surveillance approach, recommending uniform cystoscopic evaluation for all patients (Chang et al. 2016; Hall et al. 2007).

Our study design has several strengths. In aim 1, we used multilevel (mixed effects) logistic and Poisson regression models to explicitly account for the hierarchical nature of the data (patients clustered within physicians) and therefore produces correct standard errors (SEs) which do not need to be adjusted ex-post using clustered SEs. Unlike prior reports, we did not exclude patients who died or underwent total cystectomy during the exposure period (Schrag et al. 2003; Chamie et al. 2011; Chamie et al. 2012). Instead, we accounted for the different exposure time and patient-days during which each subject was observed and "at risk" for receiving surveillance. We were thus able to retain and analyze a broader and more representative cohort of NMIBC patients and minimize selection bias from only including potentially healthier patients, who had survived for at least two years and had not had their bladder removed. We compared estimates of bladder-cancer-specific and other-cause mortality using propensity-score weighted Fine-Gray competing risk regression and cause-specific Cox proportional hazards model. Reporting estimates for both disease-specific and other-cause mortality can help assess whether there might be unobserved residual confounding, which has not been accounted for by PS weighting. In aim

3, we used a patient-level simulation model to capture the heterogeneity of the NMIBC population and different risks over time of recurrence and progression to muscle-invasive disease. We also evaluated costs from both the healthcare and societal perspectives to characterize more fully the potential financial burden associated with surveillance.

Limitations

Several limitations should be considered. First, SEER-Medicare data do not collect follow-up information on patients' tumor recurrence or progression; neither do they capture underlying risks of recurrence or progression which may impact both surveillance decisions and survival outcomes. Second, we used different follow-up periods for all-cause and cause-specific mortality. While data on time of death was available through December 2016 for all-cause mortality, data on cause of death was available only until December 2013; therefore, we used a shorter follow-up period to assess bladder-cancer and other-cause mortality. Third, in aim 2, we had to further restrict the study cohort to patients who died or underwent total cystectomy during the first two years after diagnosis as we could not assign them to a surveillance cystoscopy group, which requires two years of observed exposure. Fourth, even after balancing unexposed and exposed patients on all measurable and clinically meaningful characteristics at baseline in the propensity-score weighted survival models, unadjusted confounding may still exist if unmeasured or unobserved factors, not included in the PS model, influenced treatment selection. It is likely that unmeasured time-varying characteristics such as tumor recurrence and cancer progression influenced both survival and surveillance use and contributed to our findings of a protective effect of cystoscopy on all-cause and other-cause mortality. Given the observational nature of our study and potential limitations of the propensity-score weighting methods used, we were not able to establish causal inference surveillance use and survival. While a pilot

randomized clinical trial evaluating surveillance schedules in patients with low- and intermediate-risk NMIBC is already underway (NCT02298998), it includes proportion of patients experiencing disease progression and recurrence as secondary outcome measures but does not include any survival endpoints.

Policy Implications

Despite the limitations described above, this dissertation offers a policy-relevant and timely contribution to patient-centered research on surveillance approaches following active treatment for cancer. Our findings highlight the importance of patient comorbidities, age, and functional status in surveillance evaluation for NMIBC and the need to explicitly address them in future guideline development. While the updated 2016 AUA/SUO guidelines recommend risk-stratified surveillance based on patient's underlying risk of recurrence and progression, they do not consider age, multiple comorbidities, and functional impairment. Just as overuse of surveillance among low-risk, and underuse among high-risk, patients may have undesirable consequences (Schroeck, Smith, and Shelton 2018), overuse of surveillance among older patients with functional decline and multiple comorbidities may provide little benefit and be potential harmful. The American Society of Clinical Oncology has underscored the importance of incorporating multiple chronic conditions in guideline development and implementation, cautioning that guideline recommendations that neglect this issue may not apply to, or could be harmful for, patients with multiple chronic conditions (Somerfield et al. 2016). Professional societies recommending risk-stratified surveillance based on tumor factors might also consider including age, disability status, and high comorbidity when determining appropriate frequency of surveillance for NMIBC patients.

The research also informs the decision making of providers, insurers, and patients by quantifying the value and risk-benefit trade-offs associated with different surveillance schedules for older NMIBC patients. We found that low-intensity risk-stratified surveillance over five years, with cystoscopy frequency increasing progressively with risk, was associated with lower number of recurrences detected, marginal increase in deaths and undetected recurrences, lower costs, and substantially fewer false positives, compared with a more frequent high-intensity risk-stratified approach and uniform surveillance. Findings from our study suggest that intermediate-risk patients with NMIBC may benefit from less frequent surveillance than high-risk patients.

Future Research Directions

An important extension of this research would be to assess surveillance use and outcomes in another data set that contains patient-level longitudinal information on clinical and disease characteristics. Linking clinical registry data with routine electronic health record (EHR) data will provide an opportunity to gain a more complete representation of the patient experience with surveillance over time. Similar linkages have been performed using the California Cancer Registry and EHR data from two large healthcare organizations in the same catchment area to better characterize the treatment pathways of women with breast cancer (Thompson, Kurian, and Luft 2015).

While this dissertation examined the cost-effectiveness of the 2016 AUA/SUO risk-stratified surveillance recommendations, further research is needed to evaluate the benefits and harms of different intervals of surveillance cystoscopy and outcomes in older NMIBC patients with poor functional status and multiple comorbidities. Patient preferences for different frequencies of surveillance is another important area for future studies that can provide more information on the explicit risk-benefit trade-offs patients are willing to make. While some

patients might have a strong preference for detecting recurrences as early as possible and avoiding progression at all cost, therefore willingly undergo more frequent cystoscopic evaluation, others might prefer to minimize potential false positive results, the physical discomfort, and costs associated with frequent surveillance.

Further research is also needed to validate the risk groups and classification used in the 2016 AUA/SUO risk-stratification algorithm and assess their performance as predictors of disease recurrence and progression. At present, there is no widely accepted classification of risk in NMIBC and risk-stratification algorithms vary across different guidelines and professional organizations (Sylvester et al. 2006; Babjuk et al. 2016; Chang et al. 2016; National Collaborating Centre for Cancer 2015). The EORTC risk tables (Sylvester et al. 2006) have been validated in several studies which demonstrate that the tables successfully stratify patients into risk groups for recurrence and progression, although they tend to overestimate the risk of recurrence in all risk groups and the risk of progression in high risk groups (Seo et al. 2010; Fernandez-Gomez et al. 2011; Altieri et al. 2012; T. Xu et al. 2013; Lammers et al. 2014; Hernandez et al. 2011). By contrast, the risk categories outlined in the 2016 AUA/SUO guidelines are not based a meta-analysis or original studies but on expert consensus regarding the likelihood of recurrence and progression. Investigating further the risk-predictive properties of some of the factors unique to the AUA/SUO algorithm (e.g., any BCG failure in high grade cases, any LVI, and any high grade prostatic urethral involvement) is an important next step. For example, the NICE guidelines did not include presence of LVI as a risk factor due to their assessment of low quality evidence suggesting that the presence of LVI increases the risk of recurrence, progression, and disease-specific survival in small study samples (National Collaborating Centre for Cancer 2015; Lotan et al. 2005; Kikuchi et al. 2009).

Another extension of this work is to incorporate dynamic risk groups in the cost-effectiveness model and allow patients to be re-classified at each tumor occurrence or recurrence based on their characteristics at the given time rather than at the start of the model. Gathering more evidence on the presence of PUNLMP in older adults with NMIBC and including PUNLMP as part of the low-risk stratification algorithm will also improve the risk classification. Extending the simulation model using a life-time perspective and incorporating observed real-world treatment rates for NMIBC patients who progress to muscle-invasive disease over time and become candidates for chemotherapy or immunotherapy, will provide further insights into the patient disease trajectory and care continuum over time.

APPENDIX 2.1: REVIEW OF PRIOR STUDIES EXAMINING RECEIPT OF SURVEILLANCE AMONG BLADDER CANCER PATIENTS IN THE UNITES STATES

Author (Year)	Purpose	Design/ Method	Sample/Setting	Major Variables/ Instruments	Key Study Findings	Limitations/ Comments
Schrag et al. (2003)	Examine the degree to which bladder cancer patients undergo the recommended surveillance procedures and identified patient and primary care provider characteristics associated with nonadherence to these recommendations.	Logistic regression	Patients diagnosed with superficial bladder cancer diagnosed in SEER-Medicare from 1992 through 1996 and who survived for at least 3 years after diagnosis but did not have a total cystectomy N=6,717	• Patient characteristics (tumor stage, grade, sex, age at dx, race, comorbidities, median income, year of dx, registry) • Physician characteristics (primary specialty, board certification, degree type, year of med school graduation, location of med school, sex, race, case volume, primary employer)	• The actual practice of surveillance for patients differs substantially from the standards recommended in clinical guidelines. • Only 40% of the entire cohort had an examination during all five intervals; 18.1% had low-intensity surveillance.	• No risk stratification of patients. • Analysis does not reveal whether patients, physicians, patient-physician communication, or other aspects of the health care system account for the apparent underutilization of bladder surveillance. • Did not control for measures of functional status.
Chamie et al. (2011)	Characterize practice patterns in patients with high-grade non–muscle-invasive bladder cancer in relation to established guidelines.	Mixed effects model	Patients in SEER-Medicare diagnosed with high-grade non-muscle-invasive bladder cancer between 1992 and 2002 who survived at least 2 years without undergoing definitive treatment N=4,545	• Patient characteristics (tumor stage, grade, sex, age at dx, race, comorbidities, median income, year of dx, region, education, marital status) • Facility characteristics (surgeon volume, hospital volume, institution type)	• The study found marked underuse of guideline-recommended care in this potentially curable cohort. • Unexplained provider-level factors significantly contribute to this low compliance rate. • Future studies that identify barriers and modulators of provider-level adoption of guidelines are critical to improving	• Patient preferences for surveillance and treatment strategies may have confounded their findings of significant underuse. • Did not control for measures of functional status.

					care for patients with bladder cancer.	
Hollingsworth (et al. 2010)	Understand the sources of variation in the treatment of early stage disease as well as opportunities for improving the value of cancer care	2-level linear mixed models with random intercepts	SEER-Medicare data for the years 1992 through 2002. Patients were followed using Medicare claims through December 31, 2005. N = 18,276	• Primary outcome was the initial treatment intensity received, as measured by all Medicare payments for bladder cancer incurred during the first 2 years after diagnosis • Secondary outcomes, we measured bladder cancer-specific survival using the cause-of-death field available in SEER • Patient characteristics (tumor grade, stage, age at dx, sex, socioeconomic status, comorbidities, geographic region, derived **recurrence**) • Physician characteristics (sex, type of degree, location of training, graduation year)	• Provider factors accounted for 9.2% of the variation in treatment intensity, while patient level factors accounted for 23%. • Increasing provider treatment intensity did not correlate with improved cancer-specific survival. • Although much of the variation in initial treatment intensity was determined by patient-level factors, relatively little was accounted for by commonly measured factors (e.g., patient age, comorbid status, tumor grade and stage, and subsequent disease recurrence).	• The majority of the variability in treatment intensity still resided at the patient level, and the authors' model accounted for only 23% of these between-patient differences. • Several yet unknown factors seem to influence the observed variability in treatment intensity. Although it is likely that patient preference and/or noncompliance contribute to some of this variability, other possible factors need to be explored. • Did not control for measures of functional status.
Strope et al. (2010)	Assess the relationships between clinical characteristics and treatment intensity and determine the extent to which a patient's disease risk matched with their treatment intensity	Multiple logistic regression	Patients diagnosed with early stage bladder cancer (N=24,980) between 1993 and 2002 in SEER-Medicare	• The primary outcome was patient-level treatment intensity, as measured by all Medicare payments for bladder cancer that incurred within the first 2 years after diagnosis • Patient characteristics included tumor grade,	Treatment intensity was appropriately aligned with many clinical characteristics, including age, comorbidity, tumor stage, and grade. However, treatment intensity matched disease risk for only 55% and 49% of the	• Did not use the EORTC risk tables or other accepted risk-classification algorithms to stratify patients into different risk groups • Only two risk groups defined as being at low or high

				stage, age at dx, sex, socioeconomic status, comorbidities, SEER region.	lowest and highest risk patients, respectively.	risk of bladder cancer death, not risk of recurrence or progression. • Did not control for measures of functional status.
Hollenbeck et al. (2009)	Examine associations between initial treatment intensity and subsequent patient outcomes.	• Cox proportiona l hazards model • Logistic regression	Patients diagnosed with early-stage bladder cancer from January 1, 1992, through December 31, 2002 (N = 20,713)	• The primary outcome was all-cause mortality • Secondary outcomes: bladder cancer-specific mortality; subsequent interventions (radical cystectomy, systemic chemotherapy, and radiation therapy).	• The median survival of patients was similar across all four quartiles of provider treatment intensity • Although more aggressive early treatment intensity was not associated with survival, it was associated with higher rates of major medical interventions, including radical cystectomy, systemic chemotherapy, and radiation therapy.	• Treatment intensity definition was based on bladder cancer expenditures, not specific procedures which may be problematic. Important to look at procedures individually (e.g., cystoscopy use). • Potential healthy user bias. • Did not control for measures of functional status.

APPENDIX 2.2: REVIEW OF COST-EFFECTIVENESS STUDIES EXAMINING BLADDER CANCER SURVEILLANCE STRATEGIES

Author (Year)	Country	Purpose	Interventions	Data	Methods	Key Findings	Limitations
(Lotan and Roehrborn 2002)	USA	Determine the cost-effectiveness of a plan alternating a bladder tumor marker with cystoscopy and cytology at 3-months intervals or modified care versus follow-up cystoscopy and cytology every 3 months or standard care	• cystoscopy and cytology performed every 3 months as standard care • plan alternating a bladder tumor marker with cystoscopy and cytology at 3-month intervals as modified care	Input obtained from literature review	• Model: decision tree with linear recurrence rate assumptions • Time horizon: 12 months • Main outcome: disease recurrence • Perspective: Medicare • Discount rate: N/A • SA: One-way and two-way sensitivity analyses performed	• Modified care was more cost-effective at a tumor marker cost of less than $264. • On 2-way sensitivity analysis there was no significant impact of parameters at a wide range of tumor marker costs, recurrence and progression rates.	• Linear recurrence rate • Perfect adherence to standard of care assumed • No assessment of utilities • No PSA
(Nam et al. 2000)	Canada	Compare the cost of cystoscopy and cytology (standard care) to that of urinary markers (modified care) for patients with a history of superficial bladder cancer	• cystoscopy and cytology (standard care) • urinary markers (modified care)	Patient data from a Toronto hospital	• Model: decision tree • Time horizon: 3 years • Main outcome: Diagnosed disease recurrence • Perspective: societal • Discount rate: none • SA: One-way and two-way sensitivity	Urinary marker testing for follow-up of patients with superficial bladder cancer less expensive than the standard method of cystoscopy and urinary cytology	• Societal perspective incorrectly used (indirect costs not included) • Psychosocial impact of testing and different surveillance frequencies not considered • No PSA
(Lachaine et al. 2007)	Canada	Evaluate the cost and cost-efficacy of using NMP22 compared with the standard recommended monitoring	• NMP22 • Standard recommended monitoring procedure following a transurethral	Hospital data	• Model: decision tree • Time horizon: 6 months • Main outcome: Diagnosed disease recurrence	Using NMP22 would have saved $55 per patient during the first 6 months of follow-up, resulting in a cost saving of approximately 18%.	• Short-term study horizon • Generalizability

		procedure following a transurethral resection of a bladder tumor	resection of a bladder tumor (TURBT).		• Perspective: Quebec health system • Discount rate: N/A • SA: N/A	The trade-off for using NMP22 is that some patients (8.9%) would have a 3 month delay before diagnosis of a recurrence.	
(Van Kessel et al. 2013)	Netherlands	Determined whether FGFR3 mutation analysis of voided urine samples would be cost-effective to partly replace cystoscopy in the surveillance of NMIBC patients	• Modified surveillance: FGFR3 mutation analysis of voided urine samples every 3 months, and cystoscopy at 3, 12 and 24 months • Standard surveillance: cystoscopy every 3 months • Minimal surveillance: cystoscopy at 3, 12 and 24 months	Data on 70 Dutch patients with FGFR3 positive primary tumors and a median followup of 8.8 years	• Model: Markov • Time horizon: 2 years • Main outcome: probability of no recurrence after 2 years of surveillance • Perspective: hospital • Discount rate: N/A • SA: One-way and two-way sensitivity	The total cost of surveillance after the primary tumor was lower for minimal and modified surveillance (€2,254 and €2,558, respectively) than for standard surveillance (€5,861). Results were robust to changing inputs over plausible ranges.	• No PSA • Outcomes not projected beyond 2 years • Assumptions about natural disease course to simplify the model
(De Bekker-Grob et al. 2009)	Netherlands	To determine how good microsatellite analysis (MA) markers in voided urine samples should be to make a surveillance procedure cost-effective in which	• Cystoscopy of the urinary bladder every 3 months (conventional arm) • semi-automated MA of voided urine samples to identify loss of heterozygosity every 3 months,	Data were used from a randomized trial (including 448 NMIBC patients from 10 hospitals), and from other data sources.	• Model: semi-Markov • Time horizon: 2 years • Main outcome: probability of no recurrence after 2 years of surveillance • Perspective: societal • Discount rate: N/A	The probability of being without recurrence after 2 years of surveillance was similar (86.6% conventional arm vs 86.3% test arm) with currently available MA markers (sensitivity of 58% and specificity of 73%). However, the	• No PSA • Outcomes not projected beyond 2 years Assumptions about natural disease course to simplify the model

		cystoscopy is partly replaced by MA for NMIBC patients	with a control cystoscopy at 3, 12 and 24 months(test arm).		• SA: One-way and two-way sensitivity analyses	test arm led to higher costs (€4104 vs €3433 per head).	
(Gayed, Seideman, and Lotan 2013)	USA	Assess the cost-effectiveness of using fluorescence in situ hybridization assays to determine the need for biopsy in patients with atypical cytology and equivocal or negative cystoscopy	• Biopsy • Biopsy based on FISH	Combined data from 2 large prospective studies evaluating the usefulness of fluorescence in situ hybridization in the setting of atypical cytology to detect urothelial carcinoma	• Model: decision tree • Time horizon: 2 years • Main outcome: cost per cancer detected • Perspective: hospital • Discount rate: N/A • SA: One-way sensitivity analysis	Among patients with negative cystoscopy biopsy based on fluorescence in situ hybridization resulted in costs savings of $2,241 per patient, avoiding 167 biopsies, compared to biopsy in all patients. Assuming office based biopsy, the cost savings were $216 per patient.	• Generalizability • No PSA
(Kamat et al. 2011)	USA	To assess the cost-effectiveness of using cytological evaluation, NMP22 BladderChek®, and fluorescence in situ Hybridization (FISH) UroVysion® in addition to cystoscopy in patients with a history of bladder cancer undergoing	• cystoscopy alone • cystoscopy and NMP22 • cystoscopy and FISH • cystoscopy and cytology • cystoscopy and positive NMP22 confirmed by positive FISH	Data from 200 consecutive patients with a history of bladder cancer not invading the muscle were prospectively enrolled at The University of Texas MD Anderson Cancer Center.	• Prospective clinical trial • Decision analysis not used • Patient enrollment data from August 2006 to January 2007 • Median time to first follow-up was 4.1 months after study entry	• Cystoscopy alone remains the most cost-effective strategy to detect recurrence of bladder cancer not invading the muscle. • The addition of urinary markers adds to cost, without improved detection of invasive disease.	• Generalizability (one academic cancer center) • Short-term study horizon (no long-term outcomes assessed) • No assessments of utility or the psychological impact of missing tumors (due to false-negative results) or undergoing unnecessary surgeries (due to false-positive results) when

		surveillance for recurrence.					using the different surveillance strategies.
(Zhang, Denton, and Nielsen 2013)	USA	Compare the international guidelines with alternative surveillance strategies for low-risk bladder cancer patients	• EAU guidelines for low-risk patients: cysto at 3 months; if negative, then follow-up cysto at 9 months and, subsequently, at yearly intervals for 5 years • Old AUA guidelines: no risk-stratification; cysto every 3 months in the first 2 years; every 6 months for subsequent 2–3 years; annually thereafter	Input data obtained from the published literature (EORTC risk table, CDC mortality data, other studies).	• Model: partially observable Markov model • Time horizon: lifetime • Main outcomes: QALYs; expected lifelong progression probability; lifetime number of cystoscopies. • Perspective: N/A • Discount rate: yearly • SA: One-way sensitivity analysis	• Age and comorbidity significantly affect the optimal surveillance strategy. • Results suggest that younger patients should be screened more intensively than older patients, and patients having comorbidity should be screened less intensively.	• No costs assessed in the model • No PSA

APPENDIX 3.1: DIAGNOSTIC CODES TO IDENTIFY PATIENT COHORT

SEER Code for Cancer Site	**Description**	**File**
C67.0-C67.9	Bladder	PEDSF
SEER Code for Histology	**Description**	**File**
8050-8052 8120-8124 8130-8131	Papillary carcinoma Transitional cell carcinoma Papillary transitional cell carcinoma	PEDSF
SEER AJCC-6 T Codes (2004+)	**Description**	**File**
01, 05, 10-19, 80, 81	Ta, Tis, T1, T1mic, T1a, T1a1, T1a2, T1b, T1b1, T1b2, T1c, T1 NOS, T1a NOS, T1b NOS	PEDSF
SEER AJCC-6 N Codes (2004+)	**Description**	**File**
00, 01-04, 99	N0, N0(i-), N0(i+0), N0(mol-), N0(mol+), NX	PEDSF
SEER AJCC-6 M Codes (2004+)	**Description**	**File**
00	M0	PEDSF
T value - based on AJCC 3rd (2001-2003)	**Description**	**File**
00, 01, 10, 11, 12, 13, 16, 17, 19	Tis, Ta, T1, T1a, T1b, T1c, T1a1, T1a2, T1x	PEDSF
N value - based on AJCC 3rd (2001-2003)	**Description**	**File**
00, 99	N0, NX	PEDSF
M value - based on AJCC 3rd (2001-2003)	**Description**	**File**
00	M0	PEDSF
ICD-9-CM Diagnostic Codes	**Description**	**File**
188.0 188.1 188.2 188.3 188.4 188.5 188.6 188.7 188.8	Malignant neoplasm of trigone of urinary bladder Malignant neoplasm of dome of urinary bladder Malignant neoplasm of lateral wall of urinary bladder Malignant neoplasm of anterior wall of urinary bladder Malignant neoplasm of posterior wall of urinary bladder Malignant neoplasm of bladder neck Malignant neoplasm of ureteric orifice Malignant neoplasm of urachus Malignant neoplasm of other specified sites of bladder	inpatient, outpatient, durable medical equipment, hospice, home health, or carrier-based Medicare claim

188.9	Malignant neoplasm of bladder, part unspecified	
233.7	Carcinoma in situ of bladder	
ICD-10-CM Diagnostic Codes*	**Description**	**File**
C67.0	Malignant neoplasm of trigone of urinary bladder	inpatient, outpatient, durable medical equipment, hospice, home health, or carrier-based Medicare claim
C67.1	Malignant neoplasm of dome of urinary bladder	
C67.2	Malignant neoplasm of lateral wall of urinary bladder	
C67.3	Malignant neoplasm of anterior wall of urinary bladder	
C67.4	Malignant neoplasm of posterior wall of urinary bladder	
C67.5	Malignant neoplasm of bladder neck	
C67.6	Malignant neoplasm of ureteric orifice	
C67.7	Malignant neoplasm of urachus	
C67.8	Malignant neoplasm overlapping sites of bladder	
C67.9	Malignant neoplasm of bladder, unspecified	
D09.0	Carcinoma in situ of bladder	

Note: Medicare reimbursement claims with a date of service on or after October 1, 2015 require the use of ICD-10-CM codes. The data used in for the proposed dissertation include only SEER-linked Medicare claims until December 31, 2014 which is before the date ICD-10-CM coding system was officially implemented in the United States.

APPENDIX 3.2: MEDICARE CODES USED TO IDENTIFY RELEVANT PROCEDURES

Process of care	ICD-9-CM Procedure codes (MEDPAR)	HCPCS/CPT code (Outpatient, carrier, DME)
Endoscopic surveillance		
Cystoscopy: with irrigation and evacuation of blood clots; with ureteral catheter; with brush biopsy of the ureter	57.32	52000, 52001, 52005, 52007
Cystoscopy with biopsy, fulguration	57.33	52204, 52214
Cystoscopy with insertion of radioactive substance, with or without biopsy or fulguration	57.33	52250
Cystoscopy with bladder dilation	57.33	52260, 52265
Cystoscopy with urethral dilation or urethrotomy		52270, 52275, 52276, 52277, 52281, 52282, 52283, 52285
Cystoscopy with ureteral meatotomy		52290, 52300, 52301, 52305
Cystoscopy with removal of foreign body		52310, 52315
Cystoscopy for ureteral calculus		52320, 52325, 52327, 52330, 52332, 52334
Cystoscopy with lithalopaxy		52317, 52318
Cystoscopy for ureteral stricture		52341, 52342, 52343
Cystoscopy with transurethral prostate surgery		52347, 52400, 52450, 52500, 52510, 52601, 52606, 52612, 52614, 52620, 52630, 52640, 52647, 52648, 52700
Cystoscopy with ureteroscopy	56.31, 56.33	52344, 52345, 52346, 52351, 52352, 52353, 52354, 52355
Cystoscopy, with fulguration or treatment of minor (<0.5 cm) bladder lesions, with or without biopsy	56.31, 56.33	52224
Cystoscopy through artificial stoma	57.31	
Total cystectomy	57.7, 57.71, 57.79	51570, 51575, 51580, 51585, 51590, 51595, 51596, 51597
TURBT		
Cystoscopy, with fulguration and/or resection of small (0.5 up to 2 cm), medium (2 up to 5 cm), large (≥ 5 cm) bladder tumors	57.49	52234, 52235, 52240

APPENDIX 3.3: AMERICAN JOINT COMMITTEE ON CANCER (AJCC) TNM STAGING SYSTEM FOR BLADDER CANCER (7TH ED., 2010)

T – Primary Tumor	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Ta	Noninvasive papillary carcinoma
Tis	Carcinoma in situ: "flat tumor"
T1	Tumor invades subephithelial connective tissue
T2	Tumor invades muscularis propria
pT2a	Tumor invades superficial muscularis propria (inner half)
pT2b	Tumor invades deep muscularis propria (outer half)
T3	Tumor invades perivesical tissue
pT3a	Microscopically
pT3b	Macroscopically (extravesical mass)
T4	Tumor invades any of the following: prostatic stroma, seminal vesicles, uterus, vagina, pelvic wall, abdominal wall
T4a	Tumor invades prostatic stroma, uterus, vagina
T4b	Tumor invades pelvic wall, abdominal wall
N – Regional Lymph Nodes	
NX	Lymph nodes cannot be assessed
N0	No lymph node metastasis
N1	Single regional lymph node metastasis in the true pelvis (hypogastric, obturator, external iliac, or presacral lymph node)
N2	Multiple regional lymph node metastasis in the true pelvis (hypogastric, obturator, external iliac, or presacral lymph node metastasis)
N3	Lymph node metastasis to the common iliac lymph nodes
M – Distant Metastasis	
M0	No distant metastasis
M1	Distant metastasis

Source: (Edge et al. 2010)

APPENDIX 3.4: ANATOMIC STAGE/PROGNOSTIC GROUPS

Stage	T	N	M
Stage 0a	Ta	N0	M0
Stage 0is	Tis	N0	M0
Stage I	T1	N0	M0
Stage II	T2a	N0	M0
	T2b	N0	M0
Stage III	T3a	N0	M0
	T3b	N0	M0
	T4a	N0	M0
Stage IV	T4b	N0	M0
	Any T	N1-3	M0
	Any T	Any N	M1

APPENDIX 4.1: DISTANCE TRAVELLED TO TREATING PROVIDER AMONG THOSE MEDICARE PATIENTS WITH NMIBC WHO RECEIVED CYSTOSCOPY AND HAD A VALID UNITED STATES POSTAL SERVICE ZIP CODE (N=48,708)

Distance travelled (miles)	Mean: 32.04 Median: 7.60	SD: 165 Min: 0 Max: 4,970*
Distance travelled (quartiles)	N of patients	%
Quartile 1: 0.0 – 3.3 miles	12,053	24.75
Quartile 2: 3.4 – 7.5 miles	12,207	25.06
Quartile 3: 7.6 – 17.9 miles	12,275	25.20
Quartile 4: 18.0 – 4,970.0 miles*	12,173	24.99

* Among the 15 patients who travelled >3,000 miles to their treating provider, 11 patients were from Hawaii, the other 4 patients were from the East or West coast. The greatest distance between two points on the same latitude in the US is 5,823 miles from Kure Atoll, Hawaii to Riviera Beach, Florida, therefore we did not exclude any observations as outliers.

APPENDIX 4.2: BASELINE CHARACTERISTICS STRATIFIED BY PATIENTS WHO RECEIVED ≥7 VS. <7 OR ≥4 VS. <4 CYSTOSCOPIES DURING THE FIRST TWO YEARS AFTER DIAGNOSIS

	≥7 cystoscopies			≥4 cystoscopies		
	YES	NO		YES	NO	
Variable	**N (%)**	**N (%)**	**P value**	**N (%)**	**N (%)**	**P value**
Total number of patients	18,142 (33.98)	35,243 (66.02)		37,593 (70.42)	15,792 (29.58)	
Mean age at cancer diagnosis (SD)	76.51 (6.38)	78.49 (7.23)	<0.0001	76.94 (6.59)	79.89 (7.55)	<0.0001
Age group at cancer diagnosis			<0.0001			<0.0001
66-69	2,858 (15.75)	4,390 (12.46)		5,661 (15.06)	1,587 (10.05)	
70-74	4,542 (25.04)	7,033 (19.96)		8,960 (23.83)	2,615 (16.56)	
75-79	4,900 (27.01)	8,188 (23.23)		9,759 (25.96)	3,329 (21.08)	
80-84	3,690 (20.34)	7,940 (22.53)		7,946 (21.14)	3,684 (23.33)	
≥85	2,152 (11.86)	7,692 (21.83)		5,267 (14.01)	4,577 (28.98)	
Sex			0.5689			0.8670
Male	13,697 (75.50)	26,529 (75.27)		28,319 (75.33)	11,907 (75.40)	
Female	4,445 (24.50)	8,714 (24.73)		9,274 (24.67)	3,885 (24.60)	
Race/ethnicity			<0.0001			<0.0001
White (non-Hispanic)	17,044 (93.95)	32,523 (92.28)		35,252 (93.77)	14,315 (90.65)	
Black (non-Hispanic)	441 (2.43)	1,351 (3.83)		1,008 (2.68)	784 (4.96)	
Other (including Hispanic)	657 (3.62)	1,369 (3.88)		1,333 (3.55)	693 (4.39)	
Marital status at diagnosis			<0.0001			<0.0001
Single (never married)	987 (5.44)	2,354 (6.68)		2,177 (5.79)	1,164 (7.37)	
Married/domestic partner [a]	11,933 (65.78)	20,398 (57.88)		24,014 (63.88)	8,317 (52.67)	
Other (separated, divorced, widowed)	4,002 (22.06)	10,105 (28.67)		8,840 (23.52)	5,267 (33.35)	
Unknown	1,220	2,386		2,562	1,044	

	(6.72)	(6.77)		(6.82)	(6.61)	
Charlson comorbidity index			<0.0001			<0.0001
0	13,757 (75.83)	23,876 (67.75)		27,910 (74.24)	9,723 (61.57)	
1	2,643 (14.57)	5,596 (15.88)		5,609 (14.92)	2,630 (16.65)	
2	940 (5.18)	2,686 (7.62)		2,134 (5.68)	1,492 (9.45)	
≥3	802 (4.42)	3,085 (8.75)		1,940 (5.16)	1,947 (12.33)	
Predicted Disability Status (DS)[b]			<0.0001			<0.0001
Good DS (0-2)	17,662 (97.35)	32,283 (91.60)		36,260 (96.45)	13,685 (86.66)	
Poor DS (3-4)	480 (2.65)	2,960 (8.40)		1,333 (3.55)	2,107 (13.34)	
Prior history of cancer	4,900 (27.01)	10,704 (30.37)	<0.0001	10,459 (27.82)	5,145 (32.58)	<0.0001
Secondary cancers after bladder cancer diagnosis	3,252 (17.93)	5,225 (14.83)	<0.0001	6,381 (16.97)	2,096 (13.27)	<0.0001
T classification			<0.0001			<0.0001
Ta	11,352 (62.57)	21,795 (61.84)		24,013 (63.88)	9,134 (57.84)	
Tis	1,564 (8.62)	2,432 (6.90)		2,938 (7.82)	1,058 (6.70)	
T1	5,226 (28.81)	11,016 (31.26)		10,642 (28.31)	5,600 (35.46)	
Tumor grade			0.0033			<0.0001
1—Well differentiated	2,808 (15.48)	5,667 (16.08)		6,023 (16.02)	2,452 (15.53)	
2—Moderately differentiated	5,755 (31.72)	11,543 (32.75)		12,373 (32.91)	4,925 (31.19)	
3—Poorly differentiated	3,179 (17.52)	6,169 (17.50)		6,297 (16.75)	3,051 (19.32)	
4—Undifferentiated	3,295 (18.16)	6,127 (17.39)		6,516 (17.33)	2,906 (18.40)	
Unknown	3,105 (17.11)	5,737 (16.28)		6,384 (16.98	2,458 (15.56)	
Any Medicaid state buy-in in the year prior to diagnosis	1,105 (6.09)	3,537 (10.04)	<0.0001	2,543 (6.76)	2,099 (13.29)	<0.0001
SEER region			0.0010			<0.0001
Northeast	4,623 (25.48)	9,498 (26.95)		9,638 (25.64)	4,483 (28.39)	

Midwest	2,452 (13.52)	4,507 (12.79)		5,074 (13.50)	1,885 (11.94)	
South	3,975 (21.91)	7,532 (21.37)		8,059 (21.44)	3,448 (21.83)	
West	7,092 (39.09)	13,706 (38.89)		14,822 (39.43)	5,976 (37.84)	

[a] Includes having a domestic partner (same sex or opposite sex or unregistered)
[b] Dichotomous indicator of good/poor predicted DS generated from the prediction model with no interactions using a cut-point of 0.110 (Davidoff et al. 2013).

APPENDIX 5.1: DESCRIPTIVE STATISTICS FOR AIM 2 COHORT 2 (N=41,743) STRATIFIED BY PATIENTS WHO RECEIVED ≥7 VS. <7 OR ≥4 VS. <4 CYSTOSCOPIES DURING THE FIRST TWO YEARS POST-DIAGNOSIS (BEFORE AND AFTER PROPENSITY SCORE WEIGHTING)

	Before propensity score weighting			**After propensity score weighting**			**Before propensity score weighting**			**After propensity score weighting**		
	≥7 cysto	**<7 cysto**		**≥7 cysto**	**<7 cysto**		**≥4 cysto**	**<4 cysto**		**≥4 cysto**	**<4 cysto**	
Variable	**%**	**%**	**SDM**	**%**	**%**	**SDM**	**%**	**%**	**SDM**	**%**	**%**	**SDM**
Total number of patients	17,344	24,399		17,344	24,399		34,083	7,660		34,083	7,660	
Age group at cancer diagnosis												
66-69	15.96	14.44	4.2%	15.96	14.87	3.0%	15.58	12.81	7.9%	15.58	14.67	2.6%
70-74	25.24	22.26	7.0%	25.24	23.53	4.0%	24.39	19.52	11.8 %	24.39	22.91	3.6%
75-79	27.16	24.23	6.7%	27.16	25.43	4.0%	26.12	22.42	8.7%	26.12	25.60	1.2%
80-84	20.04	21.78	4.3%	20.04	21.23	2.9%	20.69	22.68	4.8%	20.69	21.26	1.4%
≥85	11.60	17.31	16.3%	11.60	14.95	9.5%	13.22	22.58	24.6 %	13.22	15.57	6.2%
Sex												
Male	75.40	75.15	0.6%	75.40	75.32	0.2%	75.10	75.90	1.9%	75.10	75.12	0%
Female	24.60	24.85	0.6%	24.60	24.68	0.2%	24.90	24.10	1.9%	24.90	24.88	0%
Race/ethnicity												
White (non-Hispanic)	93.91	92.78	4.5%	93.91	93.26	2.6%	93.85	90.57	12.2 %	93.85	92.81	3.9%
Black (non-Hispanic)	2.44	3.48	6.1%	2.44	3.00	3.3%	2.66	4.77	11.1 %	2.66	3.24	3.1%
Other (including Hispanic)	3.64	3.75	0.5%	3.64	3.74	0.5%	3.49	4.66	5.9%	3.49	3.95	2.3%
Marital status at diagnosis												
Single (never married)	5.46	6.61	4.8%	5.46	5.94	2.0%	5.81	7.57	7.1%	5.81	6.26	1.8%
Married/domestic partner	65.93	59.90	12.5%	65.93	62.61	6.9%	64.13	54.71	19.3 %	64.13	61.86	4.7%
Other (separated, divorced, widowed)	21.91	26.10	9.8%	21.91	24.38	5.8%	23.12	29.84	15.3 %	23.12	24.74	3.7%
Unknown	6.70	7.39	2.7%	6.70	7.07	1.5%	6.93	7.87	3.6%	6.93	7.15	0.8%

Charlson comorbidity index												
0	76.26	73.10	7.3%	76.26	74.36	4.4%	75.43	69.88	12.5 %	75.43	73.94	3.4%
1	14.52	14.96	1.2%	14.52	14.75	0.7%	14.72	15.00	0.8%	14.72	14.79	0.2%
2	5.09	6.15	4.6%	5.09	5.73	2.8%	5.34	7.38	8.4%	5.34	5.86	2.1%
≥3	4.13	5.79	7.6%	4.13	5.17	4.8%	4.51	7.74	13.5 %	4.51	5.41	3.8%
Predicted Disability Status (DS)												
Poor DS	2.51	5.22	14.1%	2.51	3.96	7.5%	3.18	8.16	21.7 %	3.18	4.24	4.6%
History of any state buy-in in the year prior to diagnosis	6.05	8.71	10.2%	6.05	7.44	5.3%	6.56	12.26	19.6 %	6.56	7.94	4.7%
SEER region												
Northeast	25.57	27.31	3.9%	25.57	26.87	2.9%	25.90	29.63	8.3%	25.90	27.32	3.2%
Midwest	13.46	12.57	2.7%	13.46	13.20	0.8%	13.45	10.65	8.6%	13.45	12.87	1.8%
South	21.82	20.62		21.82			21.22	20.64	1.4%	21.22	20.67	1.4%
West	39.15	39.51	0.7%	39.15	39.51	0.2% 1	39.42	39.07	0.7%	39.42	39.14	0.6%
Prior history of cancer	26.61	27.98	3.1%	26.61	27.55	2.1%	27.05	29.03	4.4%	27.05	28.26	2.7%
Secondary cancers after bladder cancer diagnosis	17.59	14.91	7.3%	17.59	16.10	4.0%	16.41	4.30	5.9%	16.41	16.16	0.7%
T classification at diagnosis												
Ta	63.41	68.06	9.8%	63.41	66.24	6.0%	65.50	68.90	7.2%	65.50	66.24	1.6%
Tis	8.44	6.64	6.8%	8.44	7.45	3.7%	7.62	6.33	5.1%	7.62	7.51	0.4%
T1	28.16	25.30	6.5%	28.16	26.31	4.2%	26.88	24.77	4.8%	26.88	26.25	1.4%
Tumor grade												
1—Well differentiated	15.80	18.03	5.9%	15.80	17.12	3.5%	16.62	19.27	6.9%	16.62	17.21	1.5%
2—Moderately differentiated	32.15	35.82	7.8%	32.15	34.45	4.9%	33.76	36.67	6.1%	33.76	34.30	1.1%
3—Poorly differentiated	17.15	15.05	5.7%	17.15	15.76	3.8%	16.14	14.95	3.3%	16.14	15.64	1.4%
4—Undifferentiated	17.71	14.25	9.5%	17.71	15.67	5.6%	16.36	12.69	10.4 %	16.36	15.79	1.6%
Unknown	17.18	16.85	0.9%	17.18	17.00	0.5%	17.11	16.42	1.8%	17.11	17.10	0.2%

Abbr.: SDM, standardized difference in the means.

Note: Additional baseline characteristics included zip-code level education and median household income, residential status at diagnosis, surgeon volume, and year of cancer diagnosis; standardized differences in the means (SDM) after propensity score adjustment were <10% for all.

APPENDIX 6.1: DISTRIBUTION OF RISK FACTORS IN THE EORTC COHORT AND CORRESPONDING SCORES

Factor	Distribution	Recurrence Score	Progression Score
Number of tumors			
Single	57%	0	0
2 to 7	33%	3	3
≥ 8	10%	6	3
Tumor size			
< 3cm	82%	0	0
≥ 3 cm	18%	3	3
Prior recurrence rate			
Primary	55%	0	0
≤ 1 rec/year	20%	2	2
> 1 rec/year	25%	4	2
T category			
Ta	57%	0	0
T1	31%	1	4
Tis	12%	1	6
CIS			
No	96%	0	0
Yes	4%	1	6
Grade			
G1	44%	0	0
G2	45%	1	0
G3	11%	2	5
Total risk score		**0-17**	**0-23**

APPENDIX 6.2: 3-MONTH PROBABILITY OF RECURRENCE AND PROGRESSION BASED ON THE EORTC RISK TABLES

		Probability of recurrence				
Recurrence Score		**1-yr**	**2-yr**	**3-yr**	**4-yr**	**5-yr**
0	Low risk	3.98%	2.90%	2.37%	2.03%	1.84%
1-4	Intermediate risk	6.63%	5.06%	4.17%	3.45%	3.03%
5-9	High risk - lower	11.26%	8.53%	6.61%	5.42%	4.72%
10-17	High risk - upper	20.97%	14.34%	10.91%	9.03%	7.29%
		Probability of progression given recurrence				
Risk group		**1-yr**	**2-yr**	**3-yr**	**4-yr**	**5-yr**
	Low risk	1.26%	0.86%	2.82%	2.47%	2.18%
	Intermediate risk	3.78%	7.51%	8.15%	9.27%	10.18%
	High risk - lower	11.31%	12.15%	14.61%	18.65%	19.64%
	High risk - upper	21.70%	25.77%	26.84%	33.69%	40.39%
		Probability of progression				
Progression Score		**1-yr**	**2-yr**	**3-yr**	**4-yr**	**5-yr**
0	Low risk	0.05%	0.03%	0.07%	0.05%	0.04%
2-6	Intermediate risk	0.25%	0.38%	0.34%	0.32%	0.31%
7-13	High risk - lower	1.27%	1.04%	0.97%	1.01%	0.93%
14-23	High risk - upper	4.55%	3.69%	2.93%	3.04%	2.94%

Note: Annual probabilities for years 1, 2, 3, 4, and 5 derived from the EORTC calculator (Sylvester et al. 2006) were converted to 3-month probabilities as that was the cycle length of the patient-level simulation model in Aim 3.

REFERENCES

Altieri, VM, R Castellucci, P Palumbo, V Verratti, M Sut, R Olivieri, R Manco, et al. 2012. "Recurrence and Progression in Non-Muscle-Invasive Bladder Cancer Using EORTC Risk Tables." *Urol International* 89 (1): 61–66.

American Cancer Society. 2016. "Cancer Facts & Figures 2016."

Andersen, R M. 1995. "Revisiting the Behavioral Model and Access to Medical Care: Does It Matter?" *J Health Sco Behav* 36 (1): 1–10. doi:10.2307/2137284.

Austin, Peter. 2009. "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples." *Statistics in Medicine* 28 (September 2009): 3083–3107. doi:10.1002/sim.

Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46 (3): 399–424. doi:10.1080/00273171.2011.568786.

Austin, Peter C., Douglas S. Lee, and Jason P. Fine. 2016. "Introduction to the Analysis of Survival Data in the Presence of Competing Risks." *Circulation* 133 (6): 601–9. doi:10.1161/CIRCULATIONAHA.115.017719.

Austin, Peter C., Henrik Stryhn, George Leckie, and Juan Merlo. 2017. "Measures of Clustering and Heterogeneity in Multilevel Poisson Regression Analyses of Rates/Count Data." *Statistics in Medicine* 37 (4): 572–89. doi:10.1002/sim.7532.

Avritscher, Elenir B C, Catherine D. Cooksley, H. Barton Grossman, Anita L. Sabichi, Lois Hamblin, Colin P. Dinney, and Linda S. Elting. 2006. "Clinical Model of Lifetime Cost of Treating Bladder Cancer and Associated Complications." *Urology* 68 (3): 549–53. doi:10.1016/j.urology.2006.03.062.

Babjuk, Marko, Maximilian Burger, Richard Zigeuner, Shahrokh F. Shariat, Bas W G Van Rhijn, Eva Compérat, Richard J. Sylvester, et al. 2013. "EAU Guidelines on Non-Muscle-Invasive Urothelial Carcinoma of the Bladder: Update 2013." *European Urology* 64 (4): 639–53. doi:10.1016/j.eururo.2013.06.003.

———. 2016. "EAU Guidelines on Non-Muscle-Invasive Urothelial Carcinoma of the Bladder: Update 2016." *European Urology* 64 (4): 639–53. doi:10.1016/j.eururo.2013.06.003.

Babjuk, Marko, Willem Oosterlinck, Richard Sylvester, Eero Kaasinen, Andreas Böhle, Juan Palou-Redorta, and Morgan Rouprêt. 2011. "EAU Guidelines on Non-Muscle-Invasive Urothelial Carcinoma of the Bladder, the 2011 Update." *European Urology* 59 (6): 997–1008. doi:10.1016/j.eururo.2011.03.017.

Babjuk, Marko, Willem Oosterlinck, Richard Sylvester, Eero Kaasinend, Andreas Bohle, and Juan Palou-Redorta. 2008. "EAU Guidelines on Non-Muscle-Invasive Urothelial Carcinoma of the Bladder." *European Urology* 54 (2): 313–14. doi:10.1016/j.eururo.2008.04.052.

Bach, Peter B, Edward Guadagnoli, Deborah Schrag, Nicola Schussler, and L Joan. 2002. "Patient Demographic and Socioeconomic Characteristics in the SEER-Medicare Database : Applications and Limitations Source : Medical Care , Vol . 40 , No . 8 , Supplement : Use of the SEER-Medicare Data for Cancer-Related Health Services Research (Aug .,." *Medical Care* 40 (8).

Beadle, Beth M., Kai-Ping Liao, Linda S. Elting, Thomas a. Buchholz, K. Kian Ang, Adam S. Garden, and B. Ashleigh Guadagnolo. 2014. "Improved Survival Using Intensity-Modulated Radiation Therapy in Head and Neck Cancers: A SEER-Medicare Analysis." *Cancer* 120 (5): n/a-n/a. doi:10.1002/cncr.28372.

Black, Peter C., Gordon A. Brown, and Colin P.N. Dinney. 2009. "The Impact of Variant Histology on the Outcome of Bladder Cancer Treated with Curative Intent." *Urologic Oncology: Seminars and Original Investigations* 27 (1). Elsevier Inc.: 3–7. doi:10.1016/j.urolonc.2007.07.010.

Blick, Christopher G T, Sarfraz A. Nazir, Susan Mallett, Benjamin W. Turney, Natasha N. Onwu, Ian S D Roberts, Jeremy P. Crew, and Nigel C. Cowan. 2012. "Evaluation of Diagnostic Strategies for Bladder Cancer Using Computed Tomography (CT) Urography, Flexible Cystoscopy and Voided Urine Cytology: Results for 778 Patients from a Hospital Haematuria Clinic." *BJU International* 110 (1): 84–94. doi:10.1111/j.1464-410X.2011.10664.x.

BLS. 2018. "Bureau of Labor Statistics, U.S. Department of Labor, CPI-All Urban Consumers (Current Series), U.S. City Average, Medical Care, 2000-2014." Accessed July 6. https://beta.bls.gov/dataQuery/find?fq=survey:[cu]&s=popularity:D.

———. 2017. "Bureau of Labor Statistics. May 2017 National Occupational Employment and Wage Estimates United States." https://www.bls.gov/oes/tables.htm.

Bluethmann, Shirley M., Angela B. Mariotto, and Julia H. Rowland. 2016. "Anticipating the 'Silver Tsunami': Prevalence Trajectories and Comorbidity Burden among Older Cancer Survivors in the United States." *Cancer Epidemiology Biomarkers and Prevention* 25 (7): 1029–36. doi:10.1158/1055-9965.EPI-16-0133.

Botteman, Marc F, Chris L Pashos, Alberto Redaelli, Benjamin Laskin, and Robert Hauser. 2003. "The Health Economics of Bladder Cancer: A Comprehensive Review of the Published Literature." *PharmacoEconomics* 21 (18): 1315–30. doi:21183 [pii].

Bradburn, M J, T G Clark, S B Love, and D G Altman. 2003a. "Survival Analysis Part II: Multivariate Data Analysis – an Introduction to Concepts and Methods." *British Journal*

of Cancer 89 (3): 431–36. doi:10.1038/sj.bjc.6601119.

———. 2003b. "Survival Analysis Part III: Multivariate Data Analysis -- Choosing a Model and Assessing Its Adequacy and Fit." *British Journal of Cancer* 89 (4): 605–11. doi:10.1038/sj.bjc.6601120.

Brezinski, D, P H Stone, J E Muller, G H Tofler, V Davis, C Parker, L H Hartley, and E Braunwald. 1991. "Prognostic Significance of the Karnofsky Performance Status Score in Patients with Acute Myocardial Infarction: Comparison with the Left Ventricular Ejection Fraction and the Exercise Treadmill Test Performance. The MILIS Study Group." *American Heart Journal* 121 (5): 1374–81. http://www.ncbi.nlm.nih.gov/pubmed/2017970.

Briggs, Andrew, Mark Sculpher, and Karl Claxton. 2006. *Decision Modelling for Health Economic Evaluation*. Oxford university press.

Brook, RH, EA McGlynn, and PD. Cleary. 1996. "Quality of Health Care. Part 2: Measuring Quality of Care." *N Engl J Med.* 335 (13): 96.

Burger, Maximilian, James W.F. Catto, Guido Dalbagni, H. Barton Grossman, Harry Herr, Pierre Karakiewicz, Wassim Kassouf, et al. 2013. "Epidemiology and Risk Factors of Urothelial Bladder Cancer." *European Urology* 63 (2): 234–41. doi:10.1016/j.eururo.2012.07.033.

Cambier, Samantha, Richard J Sylvester, Laurence Collette, Paolo Gontero, Maurizio A Brausi, George Van Andel, Wim J Kirkels, et al. 2016. "EORTC Nomograms and Risk Groups for Predicting Recurrence , Progression , and Disease-Specific and Overall Survival in Non – Muscle-Invasive Stage Ta – T1 Urothelial Bladder Cancer Patients Treated with 1 – 3 Years of Maintenance Bacillus ´ Rin." *European Urology* 69 (1). European Association of Urology: 60–69. doi:10.1016/j.eururo.2015.06.045.

Cameron, A. Colin, and Pravin K Trivedi. 2010. *Microeconometrics Using Stata*. Revised ed. Stata Press.

Carroll, Peter R., J. Kellogg Parsons, Gerald Andriole, Robert R. Bahnson, Erik P. Castle, William J. Catalona, Douglas M. Dahl, et al. 2016. "Prostate Cancer Early Detection, Version 2.2016: Featured Updates to the NCCN Guidelines." *JNCCN Journal of the National Comprehensive Cancer Network* 14 (5): 509–19. doi:10.6004/jnccn.2016.0060.

Cesari, Matteo, Stephen B Kritchevsky, Anne B Newman, Eleanor M Simonsick, Tamara B Harris, Brenda W Penninx, Jennifer S Brach, et al. 2009. "Added Value of Physical Performance Measures in Predicting Adverse Health-Related Events: Results from the Health, Aging And Body Composition Study." *Journal of the American Geriatrics Society* 57 (2): 251–59. doi:10.1111/j.1532-5415.2008.02126.x.

Cesari, Matteo, Stephen B Kritchevsky, Brenda W H J Penninx, Barbara J Nicklas, Eleanor

M Simonsick, Anne B Newman, Frances A Tylavsky, et al. 2005. “Prognostic Value of Usual Gait Speed in Well-Functioning Older People--Results from the Health, Aging and Body Composition Study.” *Journal of the American Geriatrics Society* 53 (10): 1675–80. doi:10.1111/j.1532-5415.2005.53501.x.

Cesari, Matteo, Graziano Onder, Valentina Zamboni, Todd Manini, Ronald I Shorr, Andrea Russo, Roberto Bernabei, Marco Pahor, and Francesco Landi. 2008. “Physical Function and Self-Rated Health Status as Predictors of Mortality: Results from Longitudinal Analysis in the IlSIRENTE Study.” *BMC Geriatrics* 8 (1): 34. doi:10.1186/1471-2318-8-34.

Chamie, Karim, Eric Ballon-Landa, Timothy J. Daskivich, Jeffrey C. Bassett, Julie Lai, Jan M. Hanley, Badrinath R. Konety, Mark S. Litwin, and Christopher S. Saigal. 2015. “Treatment and Survival in Patients with Recurrent High-Risk Non-Muscle-Invasive Bladder Cancer.” *Urologic Oncology: Seminars and Original Investigations* 33 (1). Elsevier: 20.e9-20.e17. doi:10.1016/j.urolonc.2014.08.016.

Chamie, Karim, Christopher S. Saigal, Julie Lai, Jan M. Hanley, Claude M. Setodji, Badrinath R. Konety, and Mark S. Litwin. 2011. “Compliance with Guidelines for Patients with Bladder Cancer: Variation in the Delivery of Care.” *Cancer* 117 (23): 5392–5401. doi:10.1002/cncr.26198.

———. 2012. “Quality of Care in Patients with Bladder Cancer: A Case Report?” *Cancer* 118 (5): 1412–21. doi:10.1002/cncr.26402.

Chandra, A, J Shafrin, and R Dhawan. 2016. “Utility of Cancer Value Frameworks for Patients, Payers, and Physicians.” *JAMA* 315 (19): 2069–70. doi:10.1001/jama.2016.4915.

Chang, Sam, Stephen A Boorjian, Roger Chou, Peter E Clark, Siamak Daneshmand, Badrinath R Konety, Raj Pruthi, et al. 2016. “Diagnosis and Treatment of Non-Muscle Invasive Bladder Cancer: AUA/SUO Guideline.” *The Journal of Urology* 196 (June). Elsevier Ltd: 1–9. doi:10.1016/j.juro.2016.06.049.

CMS. 2012. “Discontinuance of the Unique Physician Identification Number (UPIN) Registry.”

Cohen, Jacob. 1992. “A Power Primer.” *Psychological Bulletin* 112 (1): 155–59.

Crawford, Malik, Jonathan Church, and Bradley Akin. 2016. “CPI Detailed Report Data for February 2016.” http://www.bls.gov/cpi/cpid1602.pdf.

Curran, Patrick J., and Daniel J. Bauer. 2007. “Building Path Diagrams for Multilevel Models.” *Psychological Methods* 12 (3): 283–97. doi:10.1037/1082-989X.12.3.283.

Davidoff, Amy J. 2014. “Validation of Disability Status, a Claims-Based Measure of Functional Status for Cancer Treatment and Outcomes Studies.” *Med Care* 52 (6): 500–

510. doi:10.1016/j.pestbp.2011.02.012.Investigations.

Davidoff, Amy J., Ilene H. Zuckerman, Naimish Pandya, Franklin Hendrick, Xuehua Ke, Arti Hurria, Stuart M. Lichtman, Arif Hussain, Jonathan P.Weiner, and Martin J. Edelman. 2013. "A Novel Approach to Improve Health Status Measurement in Observational Claims-Based Studies of Cancer Treatment and Outcomes." *Journal of Geriatric Oncology* 4 (2). Elsevier Inc.: 157–65. doi:10.1016/j.jgo.2012.12.005.

De Bekker-Grob, Esther W., Madelon N M Van Der Aa, Ellen C. Zwarthoff, Marinus J C Eijkemans, Bas W. Van Rhijn, Theo H. Van Der Kwast, and Ewout W. Steyerberg. 2009. "Non-Muscle-Invasive Bladder Cancer Surveillance for Which Cystoscopy Is Partly Replaced by Microsatellite Analysis of Urine: A Cost-Effective Alternative?" *BJU International* 104 (1): 41–47. doi:10.1111/j.1464-410X.2008.08323.x.

Donabedian, Avedis. 1966. "Evaluating the Quality of Medical Care." *The Milbank Quarterly* 44 (No. 3, Pt. 2): 166–203. doi:10.1111/j.1468-0009.2005.00397.x.

Drummond, Michael F, Greg L Stoddard, and George W Torrance. 2005. *Methods for the Economic Evaluation of Health Care Programmes*. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford university press.

Earle, Craig C, Ann B Nattinger, Arnold L Potosky, and Kathleen Lang. 2002. "Identifying Cancer Relapse Using SEER-Medicare Data Mallick , Mark Berger and Joan L . Warren Source : Medical Care , Vol . 40 , No . 8 , Supplement : Use of the SEER-Medicare Data for Cancer-Related Health Services Research (Aug ., 2002), Pp . IV75-IV8" 40 (8).

Edge, S, DR Byrd, CC Compton, and Et Al. 2010. *AJCC Cancer Staging Manual*. 7th ed. Springer-Verlag New York.

Feeley, Thomas W, Heidi Wied Albright, Ronald Walters, and Thomas W Burke. 2010. "A Method for Defining Value in Healthcare Using Cancer Care as a Model." *Journal of Healthcare Management* 55 (6): 399–412. http://www.ncbi.nlm.nih.gov/pubmed/21166323.

Fernandez-Gomez, Jesus, Rosario Madero, Eduardo Solsona, Miguel Unda, Luis Martinez-Piñeiro, Marcelino Gonzalez, Jose Portillo, et al. 2009. "Predicting Nonmuscle Invasive Bladder Cancer Recurrence and Progression in Patients Treated With Bacillus Calmette-Guerin: The CUETO Scoring Model." *Journal of Urology* 182 (5). Elsevier Inc.: 2195–2203. doi:10.1016/j.juro.2009.07.016.

Fernandez-Gomez, Jesus, Rosario Madero, Eduardo Solsona, Miguel Unda, Antonio Ojea, Jose Portillo, Manuel Montesinos, et al. 2011. "The EORTC Tables Overestimate the Risk of Recurrence and Progression in Patients with Non – Muscle-Invasive Bladder Cancer ´ Rin : External Validation of the Treated with Bacillus Calmette-Gue EORTC Risk Tables" 60: 423–30. doi:10.1016/j.eururo.2011.05.033.

Fine, Jason P, and Robert J Gray. 1999. “Proportional Hazards Model for the Subdistribution of a Competing Risk A Proportional Hazards Model for the Subdistribution of a Competing Risk.” *Journal of the American Statistical Association* 1459 (April 2013): 37–41.

Flaig, Thomas W., Philippe E. Spiess, Neeraj Agarwal, Rick Bangs, Stephen A. Boorjian, and Mark K. Buyyounouski. 2018. “NCCN Guidelines Version 4.2018: Bladder Cancer.” *National Comprehensive Cancer Network, Inc.*

Fleurence, Rachael L, and Christopher S Hollenbeak. 2007. “Rates and Probabilities in Economic Modelling Transformation , Translation and Appropriate Application” 25 (1): 3–6.

Garg, Tullika, Jill Nault, Ilene G Ladd, Tyler L Bogaczyk, and Sharon L Larson. 2018. “Defining Priorities to Improve Patient Experience in Non-Muscle Invasive Bladder Cancer” 4: 121–28. doi:10.3233/BLC-170138.

Garg, Tullika, Amanda J. Young, Korey A. Kost, John F. Danella, Sharon Larson, Matthew E. Nielsen, and H. Lester Kirchner. 2018. “Burden of Multiple Chronic Conditions among Patients with Urological Cancer.” *Journal of Urology* 199 (2). Elsevier Ltd: 543–50. doi:10.1016/j.juro.2017.08.005.

Garrido, Melissa M., Amy S. Kelley, Julia Paris, Katherine Roza, Diane E. Meier, R. Sean Morrison, and Melissa D. Aldridge. 2014. “Methods for Constructing and Assessing Propensity Scores.” *Health Services Research* 49 (5): 1701–20. doi:10.1111/1475-6773.12182.

Gayed, Bishoy A., Casey Seideman, and Yair Lotan. 2013. “Cost-Effectiveness of Fluorescence in Situ Hybridization in Patients with Atypical Cytology for the Detection of Urothelial Carcinoma.” *Journal of Urology* 190 (4). Elsevier Ltd: 1181–86. doi:10.1016/j.juro.2013.03.117.

Gore, John L., Mark S. Litwin, Julie Lai, Elizabeth M. Yano, Rodger Madison, Claude Setodji, John L. Adams, and Christopher S. Saigal. 2010. “Use of Radical Cystectomy for Patients with Invasive Bladder Cancer.” *Journal of the National Cancer Institute* 102 (11): 802–11. doi:10.1093/jnci/djq121.

“Greater California Registry.” 2017. https://seer.cancer.gov/registries/california.html.

Greenberg, C C, J K Wind, G J Chang, R C Chen, and D Schrag. 2013. “Stakeholder Engagement for Comparative Effectiveness Research in Cancer Care: Experience of the DEcIDE Cancer Consortium.” *J Comp Eff Res* 2 (2): 117–25. doi:10.2217/cer.12.80.

Hall, M. Craig, Sam S. Chang, Guido Dalbagni, Raj Som Pruthi, John Derek Seigne, Eila Curlee Skinner, J. Stuart Wolf, and Paul F. Schellhammer. 2007. “Guideline for the Management of Nonmuscle Invasive Bladder Cancer (Stages Ta, T1, and Tis): 2007

Update." *Journal of Urology* 178 (6): 2314–30. doi:10.1016/j.juro.2007.09.003.

Harder, Valerie, Elizabeth Stuart, and James Anthony. 2011. "Propensity Score Techniques and the Assessment of Measured Covariate Balance to Test Causal Associations in Psychological Research." *Psychol Methods* 15 (3): 8–10. doi:10.1037/a0019623.Propensity.

Haukoos, Jason S, and Roger J Lewis. 2015. "The Propensity Score." *JAMA - Journal of the American Medical Association* 314 (15): 1637–38. doi:10.1001/jama.2015.13480.

Hernandez, Virginia, E De La Pena, MD Martin, C Blazquez, FJ Diaz, and C Llorente. 2011. "External Validation and Applicability of the EORTC Risk Tables for Non-Muscle-Invasive Bladder Cancer." *World J Urol* 29: 409–14. doi:10.1007/s00345-010-0635-2.

Hollenbeck, Brent K., Zaojun Ye, Rodney L. Dunn, James E. Montie, and John D. Birkmeyer. 2009. "Provider Treatment Intensity and Outcomes for Patients with Early-Stage Bladder Cancer." *Journal of the National Cancer Institute* 101 (8): 571–80. doi:10.1093/jnci/djp039.

Hollingsworth, John M., Yun Zhang, Sarah L. Krein, Zaojun Ye, and Brent K. Hollenbeck. 2010. "Understanding the Variation in Treatment Intensity among Patients with Early Stage Bladder Cancer." *Cancer* 116 (15): 3587–94. doi:10.1002/cncr.25221.

Howard, David H., and Cary P. Gross. 2015. "Producing Evidence to Reduce Low-Value Care." *JAMA Internal Medicine* 175 (12): 1–2. doi:10.1001/jamainternmed.2015.5453.Conflict.

Hu, Chung Yuan, Yan Xing, Janice N. Cormier, and George J. Chang. 2013. "Assessing the Utility of Cancer-Registry-Processed Cause of Death in Calculating Cancer-Specific Survival." *Cancer* 119 (10): 1900–1907. doi:10.1002/cncr.27968.

Inouye, S K, P N Peduzzi, J T Robison, J S Hughes, R I Horwitz, and J Concato. 1998. "Importance of Functional Measures in Predicting Mortality among Older Hospitalized Patients." *JAMA* 279 (15): 1187–93. http://www.ncbi.nlm.nih.gov/pubmed/9555758.

Institute of Medicine. 2013. "Delivering High-Quality Cancer Care: Charting a New Course for a System in Crisis." Washington, DC.

Kamat, Ashish M., Jose A. Karam, H. Barton Grossman, A. Karim Kader, Mark Munsell, and Colin P. Dinney. 2011. "Prospective Trial to Identify Optimal Bladder Cancer Surveillance Protocol: Reducing Costs While Maximizing Sensitivity." *Bju International* 108 (7): 1123–24. doi:10.1111/j.1464-410X.2011.10344.x.

"Kentucky Registry." 2017. https://seer.cancer.gov/registries/kentucky.html.

Kikuchi, Eiji, Vitaly Margulis, Pierre I. Karakiewicz, Marco Roscigno, Shuji Mikami, Yair

Lotan, Mesut Remzi, et al. 2009. “Lymphovascular Invasion Predicts Clinical Outcomes in Patients with Node-Negative Upper Tract Urothelial Carcinoma.” *Journal of Clinical Oncology* 27 (4): 612–18. doi:10.1200/JCO.2008.17.2361.

Klabunde, Carrie N., Julie M. Legler, Joan L. Warren, Laura Mae Baldwin, and Deborah Schrag. 2007. “A Refined Comorbidity Measurement Algorithm for Claims-Based Studies of Breast, Prostate, Colorectal, and Lung Cancer Patients.” *Annals of Epidemiology* 17 (8): 584–90. doi:10.1016/j.annepidem.2007.03.011.

Klabunde, Carrie N., Arnold L. Potosky, Julie M. Legler, and Joan L. Warren. 2000. “Development of a Comorbidity Index Using Physician Claims Data.” *Journal of Clinical Epidemiology* 53 (12): 1258–67. doi:10.1016/S0895-4356(00)00256-0.

Klabunde, Carrie N., Joan L . Warren, and Julie M . Legler. 2002. “Assessing Comorbidity Using Claims Data : An Overview.” *Med Care* 40 (8): pp IV-26-IV-35.

Koo, Kevin, Lisa Zubkoff, Brenda E. Sirovich, Philip P. Goodney, Douglas J. Robertson, John D. Seigne, and Florian R. Schroeck. 2017. “The Burden of Cystoscopic Bladder Cancer Surveillance: Anxiety, Discomfort, and Patient Preferences for Decision Making.” *Urology* 108. Elsevier Inc.: 122–28. doi:10.1016/j.urology.2017.07.016.

Koroukian, Siran M., Bassam Dahman, Glenn Copeland, and Cathy J. Bradley. 2010. “The Utility of the State Buy-in Variable in the Medicare Denominator File to Identify Dually Eligible Medicare-Medicaid Beneficiaries: A Validation Study: Methods.” *Health Services Research* 45 (1): 265–82. doi:10.1111/j.1475-6773.2009.01051.x.

Kulkarni, Girish S., Shabbir M H Alibhai, Antonio Finelli, Neil E. Fleshner, Michael A S Jewett, Steven R. Lopushinsky, and Ahmed M. Bayoumi. 2009. “Cost-Effectiveness Analysis of Immediate Radical Cystectomy versus Intravesical Bacillus Calmette-Guerin Therapy for High-Risk, High-Grade (T1G3) Bladder Cancer.” *Cancer* 115 (23): 5450–59. doi:10.1002/cncr.24634.

Kulkarni, Girish S., Antonio Finelli, Neil E. Fleshner, Michael A S Jewett, Steven R. Lopushinsky, and Shabbir M H Alibhai. 2007. “Optimal Management of High-Risk T1G3 Bladder Cancer: A Decision Analysis.” *PLoS Medicine* 4 (9): 1538–49. doi:10.1371/journal.pmed.0040284.

Lammers, Rianne J M, Joan Palou, Wim P J Witjes, Maria H D Janzing-pastors, Christien T M Caris, and J Alfred Witjes. 2014. “Comparison of Expected Treatment Outcomes , Obtained Using Risk Models and International Guidelines , with Observed Treatment Outcomes in a Dutch Cohort of Patients with Non-Muscle- Invasive Bladder Cancer Treated with Intravesical Chemotherapy,” 193–201. doi:10.1111/bju.12495.

Leblanc, B, AJ Duclos, F Bénard, J Côté, L Valiquette, JM Paquin, F Mauffette, R Faucher, and JP Perreault. 1999. “Long-Term Followup of Initial Ta Grade 1 Transitional Cell Carcinoma of the Bladder.Itle.” *J Urol.* 162 (6): 1946–50.

Lee, Cheryl T., Rodney L. Dunn, Collette Ingold, James E. Montie, and David P. Wood. 2007. "Early-Stage Bladder Cancer Surveillance Does Not Improve Survival If High-Risk Patients Are Permitted to Progress to Muscle Invasion." *Urology* 69 (6): 1068–72. doi:10.1016/j.urology.2007.02.064.

Lotan, Yair, Amit Gupta, Shahrokh F. Shariat, Ganesh S. Palapattu, Amnon Vazina, Pierre I. Karakiewicz, Patrick J. Bastian, et al. 2005. "Lymphovascular Invasion Is Independently Associated with Overall Survival, Cause-Specific Survival, and Local and Distant Recurrence in Patients with Negative Lymph Nodes at Radical Cystectomy." *Journal of Clinical Oncology* 23 (27): 6533–39. doi:10.1200/JCO.2005.05.516.

Lotan, Yair, and Claus G Roehrborn. 2002. "Cost-Effectiveness of a Modified Care Protocol Substituting Bladder Tumor Markers for Cystoscopy for the Followup of Patients with Transitional Cell Carcinoma of the Bladder: A Decision Analytical Approach." *The Journal of Urology* 167 (1): 75–79.

"Louisiana Registry." 2017. https://seer.cancer.gov/registries/louisiana.html.

Lughezzani, G, M Sun, S Shariat, R T Rodolphe, C Jeldres, D Liberman, F Montorsi, P Perrotte, and P Karakiewicz. 2011. "A Population-Based Competing-Risks Analysis of the Survival of Patients Treated With Radical Cystectomy for Bladder Cancer," 103–9. doi:10.1002/cncr.25345.

Mariotto, Angela B., K. Robin Yabroff, Yongwu Shao, Eric J. Feuer, and Martin L. Brown. 2011. "Projections of the Cost of Cancer Care in the United States: 2010-2020." *Journal of the National Cancer Institute* 103 (2): 117–28. doi:10.1093/jnci/djq495.

Mayo Clinic. 2018. "Urology Tests and Procedures."

Mohler, James L., Andrew J. Armstrong, Robert R Bahnson, Anthony Victor D'Amico, Brian J. Davis, James A. Eastham, Charles A. Enke, et al. 2016. "Prostate Cancer, Version 1.2016." *Journal of the National Comprehensive Cancer Network* 14 (1): 19–30. doi:10.6004/jnccn.2016.0004.

Nam, R K, D A Redelmeier, P E Spiess, H A Sampson, Y Fradet, and M A Jewett. 2000. "Comparison of Molecular and Conventional Strategies for Followup of Superficial Bladder Cancer Using Decision Analysis." *The Journal of Urology* 163 (3): 752–57.

National Cancer Institute. Division of Cancer Control & Population. 2016. "SEER-Medicare: Number of Cases for Selected Cancers Appearing in the Data." *Patient-Centered Care & Communication.* Accessed October 3. http://healthcaredelivery.cancer.gov/seermedicare/aboutdata/cases.html.

National Collaborating Centre for Cancer. 2015. "Bladder Cancer: Diagnosis and

Management. NICE Guideline [NG2].” https://www.nice.org.uk/guidance/ng2.

Neumann, Peter J., Joshua T. Cohen, and Milton C. Weinstein. 2014. “Updating Cost-Effectiveness — The Curious Resilience of the $50,000-per-QALY Threshold.” *New England Journal of Medicine* 371 (9): 796–97. doi:10.1056/NEJMp1405158.

“New Jersey Registry.” 2017. https://seer.cancer.gov/registries/new_jersey.html.

Nielsen, Matthew E., Angela B. Smith, Anne Marie Meyer, Tzy Mey Kuo, Seth Tyree, William Y. Kim, Matthew I. Milowsky, Raj S. Pruthi, and Robert C. Millikan. 2014. “Trends in Stage-Specific Incidence Rates for Urothelial Carcinoma of the Bladder in the United States: 1988 to 2006.” *Cancer* 120 (1): 86–95. doi:10.1002/cncr.28397.

Noon, A P, P C Albertsen, F Thomas, D J Rosario, and J W F Catto. 2013. “Competing Mortality in Patients Diagnosed with Bladder Cancer : Evidence of Undertreatment in the Elderly and Female Patients,” no. November 2012. Nature Publishing Group: 1534–40. doi:10.1038/bjc.2013.106.

Nordstrom, Beth L, Jason C Simeone, Karen G Malley, Kathy H Fraeman, Zandra Klippel, Mark Durst, John H Page, and Hairong Xu. 2016. “Validation of Claims Algorithms for Progression to Metastatic Cancer in Patients with Breast, Non-Small Cell Lung, and Colorectal Cancer.” *Frontiers in Oncology* 6 (February): 18. doi:10.3389/fonc.2016.00018.

Normand, Sharon-Lise T., Mary Beth Landrum, Edward Guadagnoli, John Z. Ayanian, Thomas J. Ryan, Paul D. Cleary, and Barbara J McNeil. 2001. “Validating Recommendations for Coronary Angiography Following Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores.” *Journal of Clinical Epidemiology* 54 (4): 387–98. doi:10.1016/S0895-4356(00)00321-8.

Noyes, Katia, Eric A. Singer, and Edward M. Messing. 2008. “Cancer Care Economics of Bladder Cancer: Cost-Enhancing Factors and Possible Cost Reduction.” *Curr Opin Urol* 18 (5). doi:10.1007/s40273-013-0063-4.Modelling.

Palou, Joan, Richard J. Sylvester, Oscar Rodríguez Faba, Rubén Parada, Juan A. Peña, Ferran Algaba, and Humberto Villavicencio. 2012. “Female Gender and Carcinoma in Situ in the Prostatic Urethra Are Prognostic Factors for Recurrence, Progression, and Disease-Specific Mortality in T1G3 Bladder Cancer Patients Treated with Bacillus Calmette-Guérin.” *European Urology* 62 (1): 118–25. doi:10.1016/j.eururo.2011.10.029.

Panageas, Katherine S., Deborah Schrag, Elyn Riedel, Peter B. Bach, and Colin B. Begg. 2003. “The Effect of Clustering of Outcomes on the Association of Procedure Volume and Surgical Outcomes.” *Annals of Internal Medicine* 139 (8): 658–665+I44. doi:10.7326/0003-4819-139-8-200310210-00009.

Penninx, B W, L Ferrucci, S G Leveille, T Rantanen, M Pahor, and J M Guralnik. 2000. "Lower Extremity Performance in Nondisabled Older Persons as a Predictor of Subsequent Hospitalization." *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 55 (11): M691-7. http://www.ncbi.nlm.nih.gov/pubmed/11078100.

Potosky, A.L., G.F. Riley, J.D. Lubitz, R.M. Mentnech, and L.G. Kessler. 1993. "Potential for Cancer Related Health Services Research Using a Linked Medicare-Tumor Registry Database." *Medical Care* 31 (8): 732–48. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed3&NEWS=N&AN=8336512.

Power, Nicholas E., and Jonathan Izawa. 2016. "Comparison of Guidelines on Non-Muscle Invasive Bladder Cancer (EAU, CUA, AUA, NCCN, NICE)." *Bladder Cancer* 2 (1): 27–36. doi:10.3233/BLC-150034.

Robins, J M, M A Hernán, and B Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11 (5): 550–60. doi:10.1097/00001648-200009000-00011.

Rosenbaum, Paul R ., and Donald B . Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.

Sanders, Gillian D., Peter J. Neumann, Anirban Basu, Dan W. Brock, David Feeny, Murray Krahn, Karen M. Kuntz, et al. 2016. "Recommendations for Conduct, Methodological Practices, and Reporting of Cost-Effectiveness Analyses." *Jama* 316 (10): 1093. doi:10.1001/jama.2016.12195.

Schned, Alan R, Angeline S Andrew, Carmen J Marsit, Karl T Kelsey, Michael S Zens, and Margaret R Karagas. 2008. "Histological Classification and Stage of Newly Diagnosed Bladder Cancer in a Population-Based Study from the Northeastern United States." *Scand J Urol Nephrol.* 42 (3): 237–42. doi:10.1080/00365590801948166.Histological.

Schrag, Deborah, Lillian J Hsieh, Farhang Rabbani, Peter B Bach, Harry Herr, and Colin B Begg. 2003. "Adherence to Surveillance Among Patients With Superficial Bladder Cancer" 95 (8).

Schrag, Deborah, Nandita Mitra, Feng Xu, Farhang Rabbani, Peter B. Bach, Harry Herr, and Colin B. Begg. 2005. "Cystectomy for Muscle-Invasive Bladder Cancer: Patterns and Outcomes of Care in the Medicare Population." *Urology* 65 (6): 1118–25. doi:10.1016/j.urology.2004.12.029.

Schroeck, Florian R., Nicholas Smith, and Jeremy B. Shelton. 2018. "Implementing Risk-Aligned Bladder Cancer Surveillance Care." *Urologic Oncology: Seminars and Original Investigations*. Elsevier, 1–8. doi:10.1016/j.urolonc.2017.12.016.

"SEER-Medicare: How the SEER & Medicare Data Are Linked." 2017. https://healthcaredelivery.cancer.gov/seermedicare/overview/linked.html.

SEER. 2016. "Data Collection Answers from the CoC, NPCR, SEER Technical Workgroup. Updated June 10, 2016." https://seer.cancer.gov/registrars/data-collection.html.

"SEER Registry Groupings for Analyses." 2017. https://seer.cancer.gov/registries/terms.html.

Seo, Kyung Won, Byung Hoon Kim, Choal Hee Park, Chun Il Kim, and Hyuk Soo Chang. 2010. "Urological Oncology The Efficacy of the EORTC Scoring System and Risk Tables for the Prediction of Recurrence and Progression of Non-Muscle-Invasive Bladder Cancer after Intravesical Bacillus Calmette-Guerin Instillation," no. Appendix 2: 165–70. doi:10.4111/kju.2010.51.3.165.

Shih, Shirley L, Paul Gerrard, Richard Goldstein, Jacqueline Mix, Colleen M Ryan, Paulette Niewczyk, Lewis Kazis, et al. 2015. "Functional Status Outperforms Comorbidities in Predicting Acute Care Readmissions in Medically Complex Patients." *J Gen Intern Med* 30 (11): 1688–95. doi:10.1007/s11606-015-3350-2.

Siegel, Rebecca L, Kimberly D Miller, and Ahmedin Jemal. 2018. "Cancer Statistics, 2018." *CA Cancer J Clin* 68 (1): 7–30. doi:10.3322/caac.21442.

Silverberg, Edwin, Catherine C Boring, and Teresa S Squires. 1990. "Cancer Statistics, 1990." *CA Cancer J Clin* 40 (1): 9–26.

Snijders, Tom A.B., and Roel J. Bosker. 2012. *Multilevel Analysis: An Introduction to Basic and Applied Multilevel Analysis*. 2nd ed. SAGE.

Snijders, Tom AB, and Roel J Bosker. 2012. *Multilevel Analysis: An Introduction to Basix and Advanced Multilevel Modeling*. 2nd ed. SAGE Publications Inc.

Somerfield, M. R., K. Bohlke, G. P. Browman, N. Denduluri, K. Einhaus, D. F. Hayes, A. A. Khorana, et al. 2016. "Innovations in American Society of Clinical Oncology Practice Guideline Development." *Journal of Clinical Oncology* 34 (26). doi:10.1200/JCO.2016.68.3524.

Soria, Francesco, Marco Moschini, Stephan Korn, and Shahrokh F Shariat. 2016. "How to Optimally Manage Elderly Bladder Cancer Patients?" *Translational Andrology and Urology* 5 (5): 683–91. doi:10.21037/tau.2016.04.08.

Spiess, Philippe, Rick Bangs, Stephen A Boorjian, Mark K Buyyounouski, Peter E Clark, Tracy M Downs, Jason A Efstathiou, et al. 2017. "Bladder Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology." *J Natl Compr Canc Netw.* 15 (10): 1240–67. doi:10.6004/jnccn.2017.0156.

Stommel, Manfred, Barbara A Given, and Charles W Given. 2002. “Depression and Functional Status as Predictors of Death among Cancer Patients.” *Cancer* 94 (10): 2719–27. http://www.ncbi.nlm.nih.gov/pubmed/12173342.

Strope, Seth A., Zaojun Ye, John M. Hollingsworth, and Brent K. Hollenbeck. 2010. “Patterns of Care for Early Stage Bladder Cancer.” *Cancer* 116 (11): 2604–11. doi:10.1002/cncr.25007.

Stürmer, Til, Manisha Joshi, Robert J. Glynn, Jerry Avorn, Kenneth J. Rothman, and Sebastian Schneeweiss. 2006. “A Review of the Application of Propensity Score Methods Yielded Increasing Use, Advantages in Specific Settings, but Not Substantially Different Estimates Compared with Conventional Multivariable Methods.” *Journal of Clinical Epidemiology* 59 (5): 437–47. doi:10.1016/j.jclinepi.2005.07.004.

Stürmer, Til, Kenneth J. Rothman, and Robert J. Glynn. 2006. “Insights into Different Results from Different Causal Contrasts in the Presence of Effect-Measure Modification.” *Pharmacoepidemiology and Drug Safety* 15 (10): 698–709. doi:10.1002/pds.1231.

Suissa, Samy. 2008. “Immortal Time Bias in Pharmacoepidemiology.” *American Journal of Epidemiology* 167 (4): 492–99. doi:10.1093/aje/kwm324.

Svatek, Robert S., Brent K. Hollenbeck, Sten Holmang, Richard Lee, Simon P. Kim, Arnulf Stenzl, and Yair Lotan. 2014. “The Economics of Bladder Cancer: Costs and Considerations of Caring for This Disease.” *European Urology* 66 (2): 253–62. doi:10.1016/j.eururo.2014.01.006.

Sylvester, Richard J., Adrian P M Van Der Meijden, Willem Oosterlinck, J. Alfred Witjes, Christian Bouffioux, Louis Denis, Donald W W Newling, and Karlheinz Kurth. 2006. “Predicting Recurrence and Progression in Individual Patients with Stage Ta T1 Bladder Cancer Using EORTC Risk Tables: A Combined Analysis of 2596 Patients from Seven EORTC Trials.” *European Urology* 49 (3): 466–75. doi:10.1016/j.eururo.2005.12.031.

“Taxi Fares in the US.” 2017. https://www.taxifarefinder.com/rates.php?lang=en.

Taylor, John A, and George A Kuchel. 2009. “Bladder Cancer in the Elderly: Clinical Outcomes, Basic Mechanisms, and Future Research Direction.” *Nat Clin Pract Urol* 6 (3): 135–44. doi:10.1038/ncpuro1315.Bladder.

Thompson, Caroline A, Allison W Kurian, and Harold S Luft. 2015. “Linking Electronic Health Records to Better Understand Breast Cancer Patient Pathways Within and Between Two Health Systems” 3: 3–4.

Toossi, Mitra. 2015. “Labor Force Projections to 2024: The Labor Force Is Growing, but Slowly.” *Monthly Labor Review*, no. December: 1–36. doi:10.1525/sp.2007.54.1.23.

Van Den Bosch, Sven, and J. Alfred Witjes. 2011. “Long-Term Cancer-Specific Survival in Patients with High-Risk, Non-Muscle-Invasive Bladder Cancer and Tumour Progression: A Systematic Review.” *European Urology* 60 (3): 493–500. doi:10.1016/j.eururo.2011.05.045.

Van Kessel, Kim E M, Lucie C. Kompier, Esther W. De Bekker-Grob, Tahlita C M Zuiverloon, Yvonne Vergouwe, Ellen C. Zwarthoff, and Ewout W. Steyerberg. 2013. “FGFR3 Mutation Analysis in Voided Urine Samples to Decrease Cystoscopies and Cost in Nonmuscle Invasive Bladder Cancer Surveillance: A Comparison of 3 Strategies.” *Journal of Urology* 189 (5). Elsevier Inc.: 1676–81. doi:10.1016/j.juro.2012.11.005.

Warren, Joan L., and K. Robin Yabroff. 2015. “Challenges and Opportunities in Measuring Cancer Recurrence in the United States.” *Journal of the National Cancer Institute* 107 (8). doi:10.1093/jnci/djv134.

Warren, Joan L, Carrie N Klabunde, Deborah Schrag, Peter B Bach, and F Riley. 2002. “Overview of the SEER-Medicare Data : Content , Research Applications , and Generalizability to the United States Elderly Population Source : Medical Care , Vol . 40 , No . 8 , Supplement : Use of the SEER-Medicare Data for Cancer-Related Health Services R.” *Med Care* 40 (8).

Warren, Joan L, Angela Mariotto, Danielle Melbert, Deborah Schrag, Paul Doria-rose, David Penson, and K Robin Yabroff. 2016. “Sensitivity of Medicare Claims to Identify Cancer Recurrence in Elderly Colorectal and Breast Cancer Patients” 54 (8): 47–54.

Witjes, Johannes A. 2006. “Management of BCG Failures in Superficial Bladder Cancer: A Review.” *European Urology* 49 (5): 790–97. doi:10.1016/j.eururo.2006.01.017.

Xu, Stanley, Colleen Ross, Marsha Raebel, Susan Shetterly, Christopher Blanchette, and David Smith. 2017. “Use of Stabilized Inverse Propensity Scores as Weights to Directly Estimate Relative Risk and Its Confidence Intervals Stanley.” *Value in Health* 155 (1): 3–12. doi:10.1007/s10549-015-3663-1.Progestin.

Xu, Tianyuan, Zhaowei Zhu, Xiaohua Zhang, Xianjin Wang, Shan Zhong, Minguang Zhang, and Zhoujun Shen. 2013. “Predicting Recurrence and Progression in Chinese Patients With Nonmuscle-Invasive Bladder Cancer Using EORTC and CUETO Scoring Models.” *Urology* 82 (2). Elsevier Inc.: 387–93. doi:10.1016/j.urology.2013.04.007.

Yabroff, K. Robin, Elizabeth B. Lamont, Angela Mariotto, Joan L. Warren, Marie Topor, Angela Meekins, and Martin L. Brown. 2008. “Cost of Care for Elderly Cancer Patients in the United States.” *Journal of the National Cancer Institute* 100 (9): 630–41. doi:10.1093/jnci/djn103.

Yabroff, K. Robin, Jennifer Lund, Deanna Kepka, and Angela Mariotto. 2011. “Economic Burden of Cancer in the United States: Estimates, Projections, and Future Research.”

Cancer Epidemiology Biomarkers and Prevention 20 (10): 2006–14. doi:10.1158/1055-9965.EPI-11-0650.

Yeung, Christina, Tuan Dinh, and Joseph Lee. 2014. "The Health Economics of Bladder Cancer: An Updated Review of the Published Literature." *PharmacoEconomics* 32 (11): 1093–1104. doi:10.1007/s40273-014-0194-2.

Zhang, Yuan, Brian T Denton, and Matthew E Nielsen. 2013. "Comparison of Surveillance Strategies for Low-Risk Bladder Cancer Patients." *Med Decis Making* 33: 198–214. doi:10.1177/0272989X12465353.