ELSEVIER

Seminars in
RADIATION
ONCOLOGY

Assessing Treatment-Related Toxicity Using Administrative Data, Patient-Reported Outcomes, or Physician-Graded Toxicity: Where Is the Truth?

Michael D. Gross, BS, Bashir Al Hussein Al Awamlh, MD, and Jim C. Hu, MD, MPH

The accurate determination of the frequency and severity of treatment-related complications is vital to informing patients and clinicians in their decision-making process. In published studies, complications are assed via administrative data, patient-reported outcomes, and physician-graded toxicity, each with their strengths and limitations. Administrative data provide a vast, accessible history of patient data, but are limited in the ability to accurately capture diagnosis and causality, and are subject to differing interpretations of billing codes. Patient-reported outcomes provide direct and nuanced descriptions of both symptoms and bother; but are by definition subjective, affected by nonrespondents, and results (scores) are often difficult to interpret for patients and clinicians alike. Physician-graded toxicity is a relatively more objective measure, but relies on both clinicians fully assessing all relevant symptoms and patients accurately reporting them to the clinician. Understanding these strengths and limitations will help clinicians become more informed readers of the published literature.
Semin Radiat Oncol 29:333–337

Introduction

Accurately assessing the frequency and severity of side effects from cancer treatment is an important goal for comparative effectiveness research, and important for informing patients and physicians in their treatment decision-making process. In the published literature, three common ways of assessing side effects include analyzing secondary data sources (eg, administrative claims data), patient-reported outcomes, and physician-graded toxicity. In this article, we review the strengths and limitations of each method, using published studies in prostate cancer to serve as examples.

Secondary Data Sources

Secondary data analysis describes the use of any cohort generated through the collection and reporting of patient-specific information prior to the design of any particular research study.[1] These include administrative, claims-based or registry-based data. For administrative data like the National Surgical Quality Improvement Program (NSQIP), information on patient complications and outcomes are reported directly to a government or health care agency.[2] Patient-specific information may also be collected from health insurance billing claims. These claims-based data can be sourced from public insurers (eg, Medicare), private insurers (eg, United Health) or a mix of the two (eg, MarketScan).[3]

Utilizing secondary data sources requires careful study design, with proper capture of events, statistical methods to minimize confounding, and data reporting. Administrative data like NSQIP is generated through manual chart abstraction by trained professionals for a predetermined number of adverse events.[2] However, in claims data, researchers must use billing codes generated from patient encounters. For instance, hematuria in NSQIP is directly captured in a predetermined data field whereas using Medicare claims data, complications are indirectly accounted for by diagnosis codes (eg, gross hematuria) or procedure codes (eg, cystoscopy with fulguration). With over 13,000 Current Procedural Terminology codes and almost 72,000 International

Department of Urology, New York Presbyterian-Weill Cornell Medical College, New York, NY
Conflicts of interest: None.
Address reprint requests to Jim C. Hu, MD, MPH, Department of Urology, New York Presbyterian-Weill Cornell Medical College, 525 East 68th Street, Starr 900, New York, NY 10065. E-mail: jch9011@med.cornell.edu

https://doi.org/10.1016/j.semradonc.2019.05.007
1053-4296/

Classification of Diseases codes, selecting the correct codes is challenging.[4,5]

Advantages

Secondary data draw information from large segments of the population, allowing the study of wider patient populations than those enrolled in clinical trials. These data sets with large numbers of patients also facilitate examination of toxicity outcomes over time, accounting for the learning curve often associated with new technologies. For example, initial studies from Surveillance, Epidemiology, and End Results (SEER)-Medicare data demonstrated increased genitourinary morbidity in robot-assisted radical prostatectomy compared to the open surgical approach.[6] However, following the first SEER-Medicare study, the risks of robot-assisted radical prostatectomy were reassessed in later years, and demonstrated a progressive annual decrease in the frequency of genitourinary complications.[7] The primary genitourinary complication reported was anastomotic stricture, which decreased by almost 40% over the study period.

Another advantage of administrative data analysis is the ability to capture medical events continuously, across all health care facilities involved in a patient's care; in contrast to clinical trials where toxicities are assessed at discrete time points and only at the primary facility which enrolled the patient. There is also less patient attrition with claims-data analysis, especially Medicare data for which changes in insurance coverage are uncommon. In contrast, for studies which collect patient-reported outcomes, missing data can be a common issue that threatens study validity.

Limitations

Claims data analysis is prone to multiple potential sources of error in coding practices. Coding practices are subjective, lack guidelines, can be dependent on the specific electronic health record system used at each facility, and may be delegated to billing mangers — all of these factors contribute to inconsistent classification. For patients under 65, switching health care insurance can limit the completeness of claims-based longitudinal follow-up.

Use of procedure codes to characterize complications has varying sensitivity. In a study of SEER-Medicare patients with prostate cancer, Medicare claims were 83% sensitive and 95% specific for the presence of urethral strictures as compared to linked survey data from the Prostate Cancer Outcomes Study.[8,9] However, claims data were only 29% and 12% sensitive for incontinence and impotence, respectively. Claims data are more likely to capture severe conditions which require surgical intervention and/or those in which the patient persistently pursues the health problem with provider(s) who ultimately code the condition. In cases of low sensitivity, misses can also be due to unbilled services (bedside bladder irrigation).

The sensitivity of International Classification of Diseases codes for procedure identification also varies widely.[4] As an example, the use of endoscopy procedure codes to assess gastrointestinal toxicity following prostate cancer treatments poses significant challenges. It is difficult to determine whether "screening" endoscopies constitute routine colon cancer screening or an investigation of symptoms of rectal toxicity. One study found that eliminating codes for screening endoscopy from inclusion criteria for rectal toxicity decreases the observed incidence of gastrointestinal toxicity by almost 50%.[10] Yet, if all prostate cancer survivors who have a screening colonoscopy are counted as having rectal toxicity, this would lead to many false positives. Thus, these considerations must be taken into account when evaluating study design and interpreting outcomes.

Patient-Reported Outcomes

Patient-reported outcomes (PRO) data sources are created through the distribution of survey materials to patients. Multiple disease-specific instruments exist to quantify symptom severity and disease-specific health-related quality of life in prostate cancer, which are described in Table 1. Among the most popular are the expanded prostate cancer index and the University of California Los Angeles-Prostate Cancer Index.[11] The surveys themselves are often validated in individual patient cohorts and comparative assessments of

Table 1 Commonly Used Instruments in Prostate Cancer-Specific Quality of Life

Instrument	Number of Domains	Number of Elements	Disease Stage	Dedicated Bowel Domain	Dedicated Sexual Domain	Dedicated Urinary Domain
ESCAP-CDV	9	34	All			
EORTC QLQ-PR25	6	25	All	X	X	X
EPIC	4	50	Early	X	X	X
FACT-P	1	12	All			
PC-QoL	10	52	Early	X	X	X
PCSI	8	29	Early	X	X	X
PORPUS	1	10	All			
UCLA-PCI	6	20	Early	X	X	X

Instruments: EPIC, expanded prostate cancer index composite; EORTC QLQ-PR25, European Organization for Research and Treatment in Cancer Quality of Life Group-Prostate Cancer Module; ESCAP-CDV, Estudio sobre la Calidad de Vida en el Cáncer de Próstata; FACT-P, functional assessment of cancer therapy-prostate cancer module; PC-QoL, prostate cancer quality of life instrument; PCSI, prostate cancer symptom indices; PORPUS, patient-oriented prostate utility scale; UCLA-PCI, University of California Los Angeles-Prostate Cancer Index.

survey quality may be performed using the Evaluating Measures of Patient-Reported Outcomes tool, which includes domains of reliability, burden on the patient, and interpretability in addition to general validity.[12]

Advantages

A major advantage is that PRO data often reveal a higher prevalence (higher sensitivity) of complications than administrative claims data or physician-reported toxicity. This is likely because physicians may neglect to code all relevant diagnoses in a medical encounter, and patients may be hesitant to report all complications to their doctors. For example, a study of cancer patients privately surveyed through the SEER-MHOS found that only between 35% and 57% of those who self-reported symptoms of urinary incontinence indicated that they had discussed these issues with their doctor.[13]

PRO data are often analyzed in scales — and the granular scores facilitate statistical analysis to compare outcomes across groups and examine changes over time. These scales allow for the characterization of both functional symptoms as well as assessments of bother. Studies using patient-reported outcomes have helped establish time trends of treatment-related side effects: radical prostatectomy is associated with significant decreases in urinary and sexual function and bother which may slowly recover; whereas external beam radiation therapy causes less incontinence, but impacts bowel function and bother which may develop as late sequelae of treatment.[11] Symptom improvement that is often revealed by PRO data cannot be assessed using claims data.

Limitations

Limitations of PRO data include missing data (eg, patient nonresponse or drop-out), patient accuracy, and reproducibility. High rates of attrition should be a caution when assessing the validity of results; missing data may not be random if nonrespondents are clinically different than respondents.[14]

Patients with different baseline functional status may introduce another bias. For example, patients with no erectile function prior to treatment would not have worsening in function after treatment. Including these men in analyses of the mean decrease in function may improperly blunt treatment effects.[15]

Patient accuracy in reporting functional status is also a potential limitation; although this is likely an even bigger problem with physician-graded toxicity. First, pretreatment expectations have been shown to affect post-treatment assessment of quality of life, with those experiencing worse complications than expected reporting significantly worse quality of life.[16] Studies that require patients to recall their functional status over a prolonged period of time are subject to an additional error: in recall, patients can overestimate or underestimate their pretreatment function as compared to their actual self-reported baseline.[17] Different studies have variously shown that patient recall underestimated urinary and bowel function,[18,19] overestimated urinary function,[19] and overestimated erectile function.[18,20]

While the granular scores of PRO data facilitate statistical analysis, interpretation of these results can be difficult. To illustrate this, Table 2 summarizes several studies which directly compared the rates of urinary, sexual, and bowel toxicity between radical prostatectomy and radiation therapy.[21-25] The frequency of urinary incontinence at one year after prostatectomy across studies ranges from 15% to 44%,[21,22] and this variation is likely at least partly due to a difficulty in interpreting PRO data. For example, while questionnaires often contain several items to assess urinary incontinence (eg, frequency and severity of incontinence, and pad use) — studies often dichotomize these complex data into "yes" vs "no" incontinence; and utilizing different cutoffs can directly lead to different study conclusions. Yet, such dichotomization is necessary because without this it is impossible to communicate the meaning of PRO score differences to patients and physicians.

Physician-Graded Toxicity

The most common instrument used for physician-graded toxicity is the National Cancer Institute's Common Terminology Criteria for Adverse Events, which provides a 5-point scale for the severity of complications.[26] Physician-graded

Table 2 Selected Comparative Studies of Prostate Cancer Treatment Complications Using Patient-Reported Outcomes. Multiple Follow-Up Time Points Are Reported per Study

Study	Follow Up (years)	Radical Prostatectomy			Radiation Therapy		
		Incontinence (%)	Sexual Dysfunction (%)	Bowel Bother (%)	Incontinence (%)	Sexual Dysfunction (%)	Bowel Bother (%)
[21] *Barocas* et al. *2017*	1	15	74	3	6	72	8
	3	14	70	3	5	71	6
[22] *Schapira* et al. *2001*	1	44	89	-	8	75	-
[23] *Resnick* et al. *2013*	2	9.6	78.8	2.9	3.2	60.8	7.9
	5	13.4	75.7	4.4	4.4	71.9	5.8
	15	18.3	78.0	5.2	9.4	93.9	16.0
[24] Potosky et al. 2000	2	28.3	82.1	4.1	2.5	50.3	5.7
[25] Potosky et al. 2004	5	28.6	79.3	4.8	4.2	63.5	4.0

Table 3 Advantages and Limitations of Claims Data and Patient Report

	Advantages	Limitations
Claims and registry data	• Large sample size • Potential for continuous and long-term follow-up • Can assess trends over time and thereby account for learning curve of new technologies	• Difficult to attribute causation of complications • Retrospective data analysis subject to confounding • Prone to errors and variation in coding
Patient-reported outcomes	• Most sensitive method for capturing treatment-related toxicity • Data granularity (scores) facilitate statistical analysis and maximize statistical power to detect differences across groups • Able to assess improvements in patient symptoms over time	• Prolonged recall increases errors in patient reporting • Prone to missing data and bias introduced by non-respondents • Subjective • Results (scores) are difficult to interpret, but dichotomizing scores to facilitate interpretation introduces bias and variation
Physician-graded toxicity	• Relatively objective • CTCAE results can be compared across diseases • Complications may be directly attributed to disease or treatment	• Dependent on comprehensive assessment by physicians and accurate patient report • It is well-documented that physician-graded toxicity routinely under-captures the frequency and severity of symptoms compared to patient-report

toxicity relies upon the physician to accurately and comprehensively gather information from the patient, and the patient accurately reporting this information to the physician.

Advantages

Physician-graded toxicity is often considered an "objective" measure of treatment-related toxicity, in contrast to patent-reported outcomes which are sometimes viewed as being "subjective." Physicians are also able to infer causation in a patient's symptoms, screening out issues unrelated to treatment (such as diarrhea due to infection rather than radiotherapy).

An advantage of Common Terminology Criteria for Adverse Events is that it can be used across all diseases. Thus, a specific symptom experienced by a prostate cancer patient is graded on the same scale as the same symptom experienced by a bladder cancer patient. In contrast, many patient-reported outcomes instruments are disease-specific, which makes comparisons across studies and across cancers difficult or impossible.

Limitations

Multiple studies have consistently shown that clinicians typically report fewer symptoms overall,[27,28] of a lower severity[29,30] than patients. There are several potential reasons for this, including patient inaccurate reporting of complications (eg, patients being hesitant to tell their physicians about all of their complications), failure by the physician to attribute symptoms to the disease or treatment, as well as general limitations due to limited clinical encounter time and communication barriers. In prostate cancer treatment specifically, overall physician agreement with patient-reported toxicities range from 14% to 64%, with one study showing physicians failing to capture 81% of patient reports of sexual dysfunction.[31]

Conclusion

While secondary data sources are often the largest and most readily accessed, they are usually the least sensitive for the detection of prostate cancer treatment complications. Complications with clear and specific treatments, such as urethral stricture, are the easiest to measure; while erectile dysfunction, incontinence, and rectal toxicity require a more nuanced approach. Physician-graded toxicity is often considered an objective measure, but multiple studies have consistently shown that physicians under-capture and under-report toxicities compared to patient-reported outcomes. Patient-reported outcomes are prone to nonrandom missing data, inaccuracies due to prolonged recall, and PRO scores are often difficult to interpret clinically. Table 3 summarizes the advantages and limitations of each method for capturing treatment-related toxicity. Clinicians should be familiar with these strengths and limitations in their interpretation of published studies to help counsel patients in the treatment decision-making process.

Acknowledgment

The Frederick J. and Theresa Dow Wallace Fund of the New York Community Trust supported this research.

References

1. Szabo V, Strang VR: Secondary analysis of qualitative data. ANS Adv Nurs Sci 20:66-74, 1997
2. Khuri SF: The NSQIP: A new frontier in surgery. Surgery 138:837-843, 2005
3. Kulaylat AS, Schaefer EW, Messaris E, et al: Truven health analytics marketscan databases for clinical research in colon and rectal surgery. Clin Colon Rectal Surg 32:54-60, 2019
4. Quan H, Parsons GA, Ghali WA: Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administrative data. Med Care 42:801-809, 2004

5. International Classification of Diseases, (ICD-10-CM/PCS) Transition - Background [https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm]
6. Hu JC, Gu X, Lipsitz SR, et al: Comparative effectiveness of minimally invasive vs open radical prostatectomy. J Am Med Assoc 302:1557-1564, 2009
7. Kowalczyk KJ, Levy JM, Caplan CF, et al: Temporal national trends of minimally invasive and retropubic radical prostatectomy outcomes from 2003 to 2007: Results from the 100% Medicare sample. Eur Urol 61:803-809, 2012
8. Potosky AL, Warren JL, Riedel ER, et al: Measuring complications of cancer treatment using the SEER-Medicare data. Med Care 40: 2002(8 Suppl). IV-62
9. Potosky AL, Harlan LC, Stanford JL, et al: Prostate cancer practice patterns and quality of life: The Prostate Cancer Outcomes Study. J Natl Cancer Inst 91:1719-1724, 1999
10. Meyer A-M, Kuo T-M, Chang Y, et al: Using procedure codes to define radiation toxicity in administrative data: The devil is in the details. Med Care 55:e36-e43, 2017
11. Ávila M, Patel L, López S, et al: Patient-reported outcomes after treatment for clinically localized prostate cancer: A systematic review and meta-analysis. Cancer Treat Rev 66:23-44, 2018
12. Valderas JM, Ferrer M, Mendívil J, et al: Scientific committee on "patient-reported outcomes" of the IN: Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. Value Health 11:700-708, 2008
13. White AJ, Reeve BB, Chen RC, et al: Urinary incontinence and health-related quality of life among older Americans with and without cancer: A cross-sectional study. BMC Cancer 13:377, 2013
14. Johnson TP, Wislar JS: Response rates and nonresponse errors in surveys. J Am Med Assoc 307:1805-1806, 2012
15. Chen RC, Clark JA, Talcott JA: Individualizing quality-of-life outcomes reporting: How localized prostate cancer treatments affect patients with different levels of baseline urinary, bowel, and sexual function. J Clin Oncol 27:3916-3922, 2009
16. Maguire R, Hanly P, Drummond FJ, et al: Expecting the worst? The relationship between retrospective and prospective appraisals of illness on quality of life in prostate cancer survivors. Psychooncology 27:1237-1243, 2018
17. Litwin MS, McGuigan KA: Accuracy of recall in health-related quality-of-life assessment among men treated for prostate cancer. J Clin Oncol 17:2882-2888, 1999
18. Fransson P, Damber J-E, Widmark A: Health-related quality of life 10 years after external beam radiotherapy or watchful waiting in patients with localized prostate cancer. Scand J Urol Nephrol 43:119-126, 2009
19. Resnick MJ, Barocas DA, Morgans AK, et al: Contemporary prevalence of pretreatment urinary, sexual, hormonal, and bowel dysfunction: Defining the population at risk for harms of prostate cancer treatment. Cancer 120:1263-1271, 2014
20. Legler J, Potosky AL, Gilliland FD, et al: Validation study of retrospective recall of disease-targeted function: Results from the prostate cancer outcomes study. Med Care 38:847-857, 2000
21. Barocas DA, Alvarez J, Resnick MJ, et al: Association between radiation therapy, surgery, or observation for localized prostate cancer and patient-reported outcomes after 3 years. The J Am Med Assoc 317:1126-1140, 2017
22. Schapira MM, Lawrence WF, Katz DA, et al: Effect of treatment on quality of life among men with clinically localized prostate cancer. Med Care 39:243-253, 2001
23. Resnick MJ, Koyama T, Fan K-H, et al: Long-term functional outcomes after treatment for localized prostate cancer. N Engl J Med 368:436-445, 2013
24. Potosky AL, Legler J, Albertsen PC, et al: Health outcomes after prostatectomy or radiotherapy for prostate cancer: Results from the Prostate Cancer Outcomes Study. J Natl Cancer Inst 92:1582-1592, 2000
25. Potosky AL, Davis WW, Hoffman RM, et al: Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: The prostate cancer outcomes study. J Natl Cancer Inst 96:1358-1367, 2004
26. Atkinson TM, Ryan SJ, Bennett AV, et al: The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): A systematic review. Support Care Cancer 24:3669-3676, 2016
27. Fromme EK, Eilers KM, Mori M, et al: How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol 22:3485-3490, 2004
28. Vogelzang NJ, Breitbart W, Cella D, et al: Patient, caregiver, and oncologist perceptions of cancer-related fatigue: Results of a tripart assessment survey. The fatigue coalition. Semin Hematol 34(3 Suppl 2):4-12, 1997
29. Edited by Ingham JM, Mohamudally A, Portenoy RK: 7.7 The measurement of pain and other symptoms. In: Hanks G, Cherny NI, Christakis NA, Fallon M, Kaasa S, Portenoy RK, eds. Oxford Textbook of Palliative Medicine, Oxford University Press; 2009. Edited by
30. Patrick DL, Ferketich SL, Frame PS, et al: National Institutes of Health State-of-the-Science Conference Statement: Symptom management in cancer: Pain, depression, and fatigue, July 15-17, 2002. J Natl Cancer Inst 95:1110-1117, 2003
31. Gravis G, Marino P, Joly F, et al: Patients' self-assessment versus investigators' evaluation in a phase III trial in non-castrate metastatic prostate cancer (GETUG-AFU 15). Eur J Cancer 50:953-962, 2014