What Is the True Number Needed to Screen and Treat to Save a Life With Prostate-Specific Antigen Testing?

Stacy Loeb, Edward F. Vonesh, E. Jeffrey Metter, H. Ballentine Carter, Peter H. Gann, and William J. Catalona

See accompanying editorial on page 345 and article on page 355

A B S T R A C T

Purpose

The European Randomized Study of Screening for Prostate Cancer (ERSPC) reported a 20% mortality reduction with prostate-specific antigen (PSA) screening. However, they estimated a number needed to screen (NNS) of 1,410 and a number needed to treat (NNT) of 48 to prevent one prostate cancer death at 9 years. Although NNS and NNT are useful statistics to assess the benefits and harms of an intervention, in a survival study setting such as the ERSPC, NNS and NNT are time specific, and reporting values at one time point may lead to misinterpretation of results. Our objective was to re-examine the effect of varying follow-up times on NNS and NNT using data extrapolated from the ERSPC report.

Materials and Methods

On the basis of published ERSPC data, we modeled the cumulative hazard function using a piecewise exponential model, assuming a constant hazard of 0.0002 for the screening and control groups for years 1 to 7 of the trial and different constant rates of 0.00062 and 0.00102 for the screening and control groups, respectively, for years 8 to 12. Annualized cancer detection and drop-out rates were also approximated based on the observed number of individuals at risk in published ERSPC data.

Results

According to our model, the NNS and NNT at 9 years were 1,254 and 43, respectively. Subsequently, NNS decreased from 837 at year 10 to 503 at year 12, and NNT decreased from 29 to 18.

Conclusion

Despite the seemingly simplistic nature of estimating NNT, there is widespread misunderstanding of its pitfalls. With additional follow-up in the ERSPC, if the mortality difference continues to grow, the NNT to save a life with PSA screening will decrease.

J Clin Oncol 29:464-467. © 2010 by American Society of Clinical Oncology

INTRODUCTION

Mortality from prostate cancer (PCa) has decreased substantially in the United States, coinciding with the initiation of widespread prostate-specific antigen (PSA)–based screening. From 1994 to 2006, mortality rates declined by an average of 4% per year, the most rapid decline observed for any cancer site.¹ Mathematical models have estimated that the stage migration induced by screening likely accounts for 45% to 70% of the observed reduction in PCa mortality through 2000.² Notably, a similar decline was observed in Tyrol, Austria after the introduction of a PSA screening program, compared with the rest of the country where screening and curative treatment were uncommonly performed.³

The European Randomized Study of Screening for Prostate Cancer (ERSPC) recently reported a 20% reduction in PCa mortality and a 41% reduction in metastatic disease at diagnosis in an intent-to-screen analysis conducted after a median follow-up time of 9 years.⁴ More recently, ERSPC estimated a mortality reduction of 31% after adjustment for noncompliance in the screening arm and contamination in the control arm.⁵ However, serious concerns were raised because the original ERSPC report included estimates indicating that a large number of men would have to be screened and treated to prevent one death from PCa.⁶ The number needed to treat (NNT) is a useful statistic to assess the balance of benefits and harms of an intervention.⁷ The goal of this study is to highlight some of the pitfalls in the calculation and interpretation of the NNT statistic and, in particular, to

From the Brady Urological Institute, Johns Hopkins Medical Institutions; National Institute on Aging, Baltimore, MD; Northwestern Feinberg School of Medicine; and University of Illinois at Chicago, Chicago, IL.

Submitted May 21, 2010; accepted October 6, 2010; published online ahead of print at www.jco.org on December 28, 2010.

Supported by the Urological Research Foundation, Prostate Specialized Programs of Research Excellence Grant No. P50 CA90386-05S2, Robert H. Lurie Comprehensive Cancer Center Grant No. P30 CA60553 (W.J.C.) and the Intramural Research Program of the National Institutes of Health, National Institute on Aging (E.J.M.).

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: William J. Catalona, MD, Northwestern Feinberg School of Medicine, 675 N St Clair, Ste 20-150, Chicago, IL 60611; e-mail: wcatalona@nmff.org.

© 2010 by American Society of Clinical Oncology

0732-183X/11/2904-464/\$20.00

DOI: 10.1200/JCO.2010.30.6373

provide revised estimates of the NNT from the ERSPC trial accounting for the important effects of longer follow-up time.

Whether or not one accepts that PSA screening has a mortality benefit or at least reduces the incidence of metastatic disease, it must be acknowledged that screening programs engender costs at both the individual and societal level.⁸ Central to the debate over PSA screening are concerns regarding the diagnosis and treatment of tumors that may not cause harm.⁹ In the ERSPC trial, Schroder et al⁴ used the difference in cumulative mortality between the screening and control arms and the excess incidence of PCa in the screening arm to estimate a NNT of 48 to prevent one PCa death after a median follow-up time of 9 years. Because not every patient diagnosed with PCa will require treatment, NNT can be described more accurately in this context as the number needed to diagnose.¹⁰ The number needed to screen (NNS), which is simply the reciprocal of the absolute difference in cumulative mortality, was initially reported by ERSPC as 1,410 at the 9-year follow-up mark. This number can be reinterpreted as the number needed to be offered screening. The NNS was 1,068 when screening arm assignees who never underwent any screening were excluded.

Previous authors noted that the NNT statistic frequently has been used incorrectly in clinical trial reports in leading journals.^{11,12} NNT is easily understood when referring to proportions of patients assigned to each group at baseline but becomes more complex when dealing with differences in time-to-event data or event rates, which are based on actual person-time of observation. First, when rates, rather than proportions, are used as the basis for estimating NNT in the context of mortality, the NNT represents the amount of person-time (usually person-years), not the number of persons, that must be treated to prevent one death. Although this approach has been advocated as a way of standardizing the observation period and thus dealing with trials that have long and varying follow-up times for patients, the results are less intuitively appealing to clinicians, and their validity depends on the assumption that risk changes at a constant rate over time.¹²⁻¹⁴ Second, in almost all long-term trials such as ERSPC, some participants are removed from observation (ie, censored as a result of

death or loss to follow-up) at varying points during follow-up, and the rates of censoring can also vary between treatment groups. Ignoring censoring, particularly differential censoring, can distort estimates of NNT that are based on simple proportions. This risk of distortion can be mitigated by instead calculating NNT based on the survival curves (or equivalent cumulative hazard functions [CHF]) for each treatment group derived from common statistics such as the Kaplan-Meier or Nelsen-Aalen estimators, which account for variations in follow-up time among patients.¹⁵ Schroder et al⁴ appear to have used the Nelsen-Aalen CHF in estimating NNS and NNT. Because recalculation of the NNS and NNT using simple proportions based on the number of patients at baseline and the number of PCa deaths in each group yields nearly identical results, we assume that the pattern of censoring was approximately equivalent in each arm of the trial.

In the current analysis, however, we focused on another concern that we believe has a major effect on interpretation of NNT and NNS in the ERSPC, namely the fact that these statistics are time specific and will change as the risks for the treatment groups either converge or diverge over time. By extracting hazard rate estimates from the authors' Nelsen-Aalen curves and applying those rates in an appropriate model, we calculated predicted NNS and NNT estimates for different periods of follow-up.

MATERIALS AND METHODS

We modeled the CHF of PCa-specific mortality for each treatment group using a piecewise exponential (PWE) model. The PWE model is a widely used approach to survival analysis that is particularly suited for situations involving nonproportional hazards (ie, those such as ERSPC where the relative hazards for PCa death and, therefore, the absolute mortality differences change over time).^{16,17} PWE models incorporate covariates into an actuarial life-table approach to survival analysis in much the way the Cox model incorporates covariates into a Kaplan-Meier approach. Unlike the Cox model, which does not specify any baseline hazard rate, the PWE model divides follow-up time into discrete, nonoverlapping intervals. The baseline hazard (ie, excluding the

Fig 1. Modeled cumulative hazard functions assuming a piecewise exponential model. NNT, number needed to treat; NNS, number needed to screen; CHR, cumulative hazard ratio; HR, hazard ratio.

effects of covariates) can vary from one interval to the next but remains constant within the interval. The PWE and Cox models have been shown to yield nearly equivalent results for estimating covariate effects in many situations. However, the PWE model allows one to make predictions for individual patients based on covariate histories and, more to the point here, allows flexibility in defining the shape of the hazard function over time.^{16,18} This flexibility is important in situations such as PCa screening trials where the delayed emergence of a mortality benefit can be expected.

In our model, we assumed a constant hazard of 0.0002 for both the screening and control groups for years 1 to 7 of the trial. This is based on assuming the CHF to be 0.001 at 5 years based directly on the estimated CHF shown in Figure 2 of Schroder et al.⁴ Similarly, for years 8 to 12 of the trial, we assumed different constant rates of 0.00062 for the screening group (assuming a CHF of 0.0045 at 12 years) and 0.00102 for the control group (assuming a CHF of 0.0065 at 12 years), all based directly on Figure 2 of Schroder et al.⁴ Given this nonproportional hazards assumption, we computed PCa-free survival and cumulative hazard ratios over time as a function of the CHF. Annualized cancer detection and drop-out rates were also approximated based on the observed number of individuals at risk in published ERSPC data.⁴

RESULTS

Figure 1 compares the modeled CHFs to published data from the ERSPC. According to our model, the NNS and NNT at 9 years were 1,254 and 43, respectively (Table 1); these numbers are close to the published figures of 1,410 and 48, respectively. Our model also corresponds to a cumulative hazard ratio of 0.77, similar to the crude hazard ratio of 0.80 from the ERSPC report. Subsequently, the NNS decreased from 837 at year 10 to 503 at year 12, and the NNT decreased from 29 at year 10 to 18 at year 12, an estimate that is similar to the one determined by Welch et al⁹ using population data from the Surveillance, Epidemiology, and End Results program and by Bill-Axelsson et al¹⁹ based on a randomized trial of surgery versus no treatment for PCa. Finally, Hugosson et al¹⁰ recently reported results from the Goteborg PCa screening trial, which was designed independently but included a subset of participants from the ERSPC. Using

Table 1. Modeled Results Assuming a Piecewise Exponential Model

Year and Group	Hazard Function	Constant Hazard Rate	Survival	Dropout Rate	No. of Patients at Risk	NNS	NNT	Hazard Ratio	Cumulative Hazard Ratio
0									
Control	0.00000	0.00020	1.00000	0.00000	89,353	—	—	1.00	—
Screening	0.00000	0.00020	1.00000	0.00000	72,890	—	—	1.00	—
1									
Control	0.00020	0.00020	0.99980	0.02186	87,421	—	—	1.00	1.00
Screening	0.00020	0.00020	0.99980	0.02267	71,256	—	—	1.00	1.00
2									
Control	0.00040	0.00020	0.99960	0.02186	85,530	—	—	1.00	1.00
Screening	0.00040	0.00020	0.99960	0.02267	69,659	—	—	1.00	1.00
3									
Control	0.00060	0.00020	0.99940	0.02186	83,681	—	—	1.00	1.00
Screening	0.00060	0.00020	0.99940	0.02267	68,097	—	—	1.00	1.00
4									
Control	0.00080	0.00020	0.99920	0.02186	81,871	—	—	1.00	1.00
Screening	0.00080	0.00020	0.99920	0.02267	66,570	—	—	1.00	1.00
5									
Control	0.00100	0.00020	0.99900	0.02186	80,101	—	—	1.00	1.00
Screening	0.00100	0.00020	0.99900	0.02267	65,078	—	—	1.00	1.00
6									
Control	0.00120	0.00020	0.99880	0.04277	76,747	—	—	1.00	1.00
Screening	0.00120	0.00020	0.99880	0.04986	61,913	—	—	1.00	1.00
7									
Control	0.00140	0.00020	0.99860	0.04277	73,534	—	—	1.00	1.00
Screening	0.00140	0.00020	0.99860	0.04986	58,902	—	—	1.00	1.00
8									
Control	0.00242	0.00102	0.99758	0.37661	50,458	—	—	1.00	1.00
Screening	0.00202	0.00062	0.99798	0.35528	41,289	2,506	86	0.61	0.83
9									
Control	0.00344	0.00102	0.99657	0.37661	34,623	—	—	1.00	1.00
Screening	0.00264	0.00062	0.99736	0.35528	28,943	1,254	43	0.61	0.77
10									
Control	0.00446	0.00102	0.99555	0.37661	23,758	—	—	1.00	1.00
Screening	0.00326	0.00062	0.99675	0.35528	20,288	837	29	0.61	0.73
11									
Control	0.00548	0.00102	0.99453	0.37661	16,302	—	—	1.00	1.00
Screening	0.00388	0.00062	0.99613	0.35528	14,221	628	22	0.61	0.71
12									
Control	0.00650	0.00102	0.99352	0.37661	11,186	—	—	1.00	1.00
Screening	0.00450	0.00062	0.99551	0.35528	9,969	503	18	0.61	0.69

Abbreviations: NNS, number needed to screen; NNT, number needed to treat.

data from extended follow-up, these investigators calculated an NNS of 293 and NNT of 12 to prevent one PCa death at a median follow-up time of 14 years, suggesting that the estimates from our PWE model are highly plausible. We note that the NNS and NNT estimates from the Göteborg trial,¹⁰ unlike the ERSPC results, seem to have been based on simple proportions and may have been overestimated. The NNS calculated using the inverse of the difference (0.40%) in the Kaplan-Meier cumulative risk of PCa death is 250.

DISCUSSION

Overall, our results demonstrate that NNS and NNT are highly sensitive to the time-dependent effects of the screening intervention on PCa mortality. Accordingly, estimates of NNS and NNT at a single time point during a survival study may be misleading. In addition to their dependence on time of follow-up and changes in the slopes of the hazard functions, NNS and NNT estimates from an intent-to-treat analysis may also be influenced by other features of a screening study, such as noncompliance and contamination. It is clear, based on both the CHF reported by Schroder et al⁴ and the estimated CHF using our PWE model (Fig 1), that the hazard rates for PCa mortality are not proportional over time and that there is a sharp increase in PCa-related deaths after 7 years that must be accounted for when estimating NNS and NNT as a function of time. Indeed, because of the long natural history of PCa, a follow-up time of more than 10 years is necessary to evaluate cancer-specific mortality.

Despite the seemingly simplistic nature of estimating NNT, there is widespread misunderstanding of its pitfalls among the medical community, the media, and the general public. Specifically, in the setting of a survival study such as the ERSPC, quoting one set of values for NNS and NNT at a single time point may be misleading. With additional follow-up in the ERSPC, the mortality difference between

the screening and control arms will likely continue to grow, thus leading to further decreases in the NNT estimates.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment or Leadership Position: None **Consultant or Advisory Role:** William J. Catalona, Beckman Coulter (U), Ohmx (U) **Stock Ownership:** None **Honoraria:** William J. Catalona, Beckman Coulter, GlaxoSmithKline **Research Funding:** William J. Catalona, Beckman Coulter **Expert Testimony:** None **Other Remuneration:** None

AUTHOR CONTRIBUTIONS

Conception and design: Stacy Loeb, Edward F. Vonesh, E. Jeffrey Metter, H. Ballentine Carter, Peter H. Gann, William J. Catalona **Administrative support:** Edward F. Vonesh, E. Jeffrey Metter, H. Ballentine Carter, William J. Catalona **Provision of study materials or patients:** Stacy Loeb, Edward F. Vonesh, William J. Catalona **Collection and assembly of data:** Stacy Loeb, Edward F. Vonesh, E. Jeffrey Metter, H. Ballentine Carter, Peter H. Gann, William J. Catalona **Data analysis and interpretation:** Stacy Loeb, Edward F. Vonesh, E. Jeffrey Metter, H. Ballentine Carter, Peter H. Gann, William J. Catalona **Manuscript writing:** Stacy Loeb, Edward F. Vonesh, E. Jeffrey Metter, H. Ballentine Carter, Peter H. Gann, William J. Catalona **Final approval of manuscript:** Stacy Loeb, Edward F. Vonesh, E. Jeffrey Metter, H. Ballentine Carter, Peter H. Gann, William J. Catalona

REFERENCES

1. National Cancer Institute: Surveillance, Epidemiology and End Results (SEER). <http://www.seer.cancer.gov>
2. Etzioni R, Tsodikov A, Mariotto A, et al: Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes Control* 19:175-181, 2008
3. Bartsch G, Horninger W, Klocker H, et al: Tyrol Prostate Cancer Demonstration Project: Early detection, treatment, outcome, incidence and mortality. *BJU Int* 101:809-816, 2008
4. Schröder FH, Hugosson J, Roobol MJ, et al: Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 360:1320-1328, 2009
5. Roobol MJ, Kerkhof M, Schröder FH, et al: Prostate cancer mortality reduction by prostate-specific antigen-based screening adjusted for non-attendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC). *Eur Urol* 56:584-591, 2009
6. Barry MJ: Screening for prostate cancer: The controversy that refuses to die. *N Engl J Med* 360:1351-1354, 2009
7. Laupacis A, Sackett DL, Roberts RS: An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 318:1728-1733, 1988
8. Welch HG: Should I Be Tested for Cancer? Maybe Not and Here's Why. Berkeley, CA, University of California Press, 2004
9. Welch HG, Albertsen PC: Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. *J Natl Cancer Inst* 101:1325-1329, 2009
10. Hugosson J, Carlsson S, Aus G, et al: Mortality results from the Göteborg randomised population-based prostate-cancer screening trial. *Lancet Oncol* 11:725-732, 2010
11. Suissa S: Calculation of number needed to treat. *N Engl J Med* 361:424-425, 2009
12. Hildebrandt M, Vervölgyi E, Bender R: Calculation of NNTs in RCTs with time-to-event outcomes: A literature review. *BMC Med Res Methodol* 9:21, 2009
13. Lubsen J, Hoes A, Grobbee D: Implications of trial results: The potentially misleading notions of number needed to treat and average duration of life gained. *Lancet* 356:1757-1759, 2000
14. Mayne TJ, Whalen E, Vu A: Annualized was found better than absolute risk reduction in the calculation of number needed to treat in chronic conditions. *J Clin Epidemiol* 59:217-223, 2006
15. Altman DG, Andersen PK: Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 319:1492-1495, 1999
16. Vonesh E, Schaubel DE, Hao W, et al: Statistical methods for comparing mortality among ESRD patients: Examples of regional/international variations. *Kidney Int Suppl* 54:S19-S27, 2000 (suppl)
17. Holford TR: The analysis of rates and of survivorship using log-linear models. *Biometrics* 36:299-305, 1980
18. Gann PH, Fought A, Deaton R, et al: Risk factors for prostate cancer detection after a negative biopsy: A novel multivariable longitudinal approach. *J Clin Oncol* 28:1714-1720, 2010
19. Bill-Axelsson A, Holmberg L, Filén F, et al: Radical prostatectomy versus watchful waiting in localized prostate cancer: The Scandinavian Prostate Cancer Group-4 randomized trial. *J Natl Cancer Inst* 100:1144-1154, 2008