

Research article

Open Access

## Estimating the cumulative risk of false positive cancer screenings

Stuart G Baker\*<sup>1</sup>, Diane Erwin<sup>2</sup> and Barnett S Kramer<sup>3</sup>

Address: <sup>1</sup>Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, Maryland, U.S.A, <sup>2</sup>Information Management Services, Inc., Rockville, Maryland, USA and <sup>3</sup>Office of Disease Prevention, National Institutes of Health, Bethesda, Maryland, USA

Email: Stuart G Baker\* - sb16i@nih.gov; Diane Erwin - d.erwin@btinternet.com; Barnett S Kramer - KramerB@OD.NIH.GOV

\* Corresponding author

Published: 03 July 2003

Received: 11 February 2003

BMC Medical Research Methodology 2003, 3:11

Accepted: 03 July 2003

This article is available from: <http://www.biomedcentral.com/1471-2288/3/11>

© 2003 Baker et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** When evaluating cancer screening it is important to estimate the cumulative risk of false positives from periodic screening. Because the data typically come from studies in which the number of screenings varies by subject, estimation must take into account dropouts. A previous approach to estimate the probability of at least one false positive in  $n$  screenings unrealistically assumed that the probability of dropout does not depend on prior false positives.

**Method:** By redefining the random variables, we obviate the unrealistic dropout assumption. We also propose a relatively simple logistic regression and extend estimation to the expected number of false positives in  $n$  screenings.

**Results:** We illustrate our methodology using data from women ages 40 to 64 who received up to four annual breast cancer screenings in the Health Insurance Program of Greater New York study, which began in 1963. Covariates were age, time since previous screening, screening number, and whether or not a previous false positive occurred. Defining a false positive as an unnecessary biopsy, the only statistically significant covariate was whether or not a previous false positive occurred. Because the effect of screening number was not statistically significant, extrapolation beyond 4 screenings was reasonable. The estimated mean number of unnecessary biopsies in 10 years per woman screened is .11 with 95% confidence interval of (.10, .12). Defining a false positive as an unnecessary work-up, all the covariates were statistically significant and the estimated mean number of unnecessary work-ups in 4 years per woman screened is .34 with 95% confidence interval (.32, .36).

**Conclusion:** Using data from multiple cancer screenings with dropouts, and allowing dropout to depend on previous history of false positives, we propose a logistic regression model to estimate both the probability of at least one false positive and the expected number of false positives associated with  $n$  cancer screenings. The methodology can be used for both informed decision making at the individual level, as well as planning of health services.

### Background

When evaluating cancer screening, it is important to estimate both the benefits and harms. The major benefit is

the reduction in mortality from the cancer that is the object of the screening [1]. A frequent harm is a false positive (FP) screening outcome. Although there are also risks

from the screening procedure, such as perforation of the colon in colorectal cancer screening using endoscopy, diagnosis of medically unimportant "cancers" (overdiagnosis) and burdens due to time lost and travel to the screening clinic, these will not be discussed here. A false positive (FP) can be defined either narrowly as an unnecessary biopsy (i.e. a biopsy that does not detect cancer) or broadly as unnecessary additional work-up (i.e. an additional work-up that does not detect cancer).

Gelfand and Wang (GW) [2-4] proposed methodology for estimating the probability of a least one FP in  $n$  screenings. We bolster and extend the methodology in GW. First, we show that an unrealistic assumption of GW is unnecessary. GW thought that they needed to assume that dropping out (either by loss-to-follow-up or refusing additional screenings) is independent of the prior history of false positives. This assumption is unrealistic [5] and makes their approach untenable. We show that by reformulating the problem with different random variables, one can obtain essentially the same result without the unrealistic assumption. Second we estimate an additional quantity to that estimated by GW. GW only estimated the probability of at least one false positive in  $n$  screenings. To better quantify the cumulative burden of false positives, we also estimated the expected number of false positives in  $n$  screenings. Third, to simplify computations for some data sets, we introduced a logistic regression model.

**Methods**

**Obviating the Unrealistic Dropout Assumption**

The proof obviating the assumption that dropout does not depend on previous FP's is technical and deferred to the Appendix (see Additional file: 1). However the main idea can be readily summarized. Unlike GW who estimate the probability of at least one false positive among  $n$  screenings that have occurred, we estimate the probability of a least one false positive *if* there were  $n$  screenings regardless of whether or not they occurred. This seemingly slight modification of the definition makes a large difference in the mathematical derivation, which in turn obviates the unrealistic assumption. Our much less restrictive assumption is that dropout does not depend on future false positives. This revised formulation has a parallel in Kaplan-Meier estimation that requires only that censoring not depend on future outcomes [6] and in discrete-time censoring models that are formulated as missing-data selection models [7].

**Logistic Regression**

To estimate parameters, GW used a Bayesian approach with a proportional hazards model. For the clinically oriented reader, the computations can be difficult. As an alternative we propose a relatively simple logistic regression models that is appropriate for some data sets (with

further elaboration in the Discussion). Our approach requires fitting two logistic regressions. The first logistic regression models the probability of FP on the first screening as a function of age. We let  $i$  index age interval at screening, where  $i = 1, 2, 3, 4, 5$  corresponds to ages 40-44, 45-49, 50-54, 55-59, 60-64, respectively. The logistic regression can be written as

$$\text{logit}(pr(FP|i;\alpha)) = \alpha_0 + \alpha_{age(i)}, \quad (1)$$

where  $\alpha_{age(1)} = 0$  because we have constant term  $\alpha_0$ . The data are a table of counts for age categories cross-classified by FP outcome (yes or no). See supplemental file.

The second logistic regression models the probability of FP on a screening after the first as a function of age at screening, time since the last screening, the number of the screening, and whether or not there was a previous FP. To obtain a parsimonious model we have made two simplifications. First we use screening number rather than chronological time. For example, in one subject screening might occur at times 0, 1, and 3, and in another subject, screening might occur at times 0, 1, 2. In terms of the model, both subjects have three screenings indexed by  $t = 1, 2, 3$ . This has the advantage of reducing the number of missing-data patterns as there are, by definition, no missing "between" screenings. To model various patterns of missingness would require a much more complicated formulation, which is not warranted given the sparse data. To account for different intervals between screenings we included time since the previous screenings as a covariate (which parallels the GW formulation). Our second simplification is using an indicator of previous FP rather than a more detailed history of FP because there are too few data to adequately fit a model that conditions on various prior patterns of FP.

As before we let  $i$  denote age interval at screening. We also let  $j$  denote time since the last screening, where  $j = 1, 2, 3, 4$  corresponds to 9-12 months, 13-15 months, 16-18 months, and greater than 18 months, respectively. We let  $k$  denote whether or not there was a previous false positive, where  $k = 0, 1$  corresponds to no and yes, respectively. The logistic regression is

$$\text{logit}(pr(FP|i, j, t, k;\beta)) = \beta_0 + \beta_{age(i)} + \beta_{time(j)} + \beta_{screen(t)} + \beta_{FP(k)}. \quad (2)$$

where  $\beta_{age(1)} = \beta_{time(1)} = \beta_{screen(1)} = \beta_{FP(1)} = 0$ . The data are counts for a cross classification of age interval, time interval, screening number, an indicator of previous FP, and the FP outcome. See Additional file: 2. The model in (2) represents a standard application of logistic regression to survival analysis [8,9].

**Estimating Cumulative Risk**

We use the parameter estimates from the logistic regression to estimate the cumulative risk of an FP. Let

$$a_{1i} = pr(Y = 1; \hat{\alpha}) = \frac{e^{\hat{\alpha}_0 + \hat{\alpha}_{age(i)}}}{1 + e^{\hat{\alpha}_0 + \hat{\alpha}_{age(i)}}}, \tag{3}$$

$$a_{ij} = pr(Y = 1 | i, j, k, 0; \hat{\beta}) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_{age(i)} + \hat{\beta}_{time(j)} + \hat{\beta}_{screen(t)} + \hat{\beta}_{FP(0)}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_{age(i)} + \hat{\beta}_{time(j)} + \hat{\beta}_{screen(t)} + \hat{\beta}_{FP(1)}}}, \tag{4}$$

$$b_{ij} = pr(Y = 1 | i, j, k, 1; \hat{\beta}) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_{age(i)} + \hat{\beta}_{time(j)} + \hat{\beta}_{screen(t)} + \hat{\beta}_{FP(1)}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_{age(i)} + \hat{\beta}_{time(j)} + \hat{\beta}_{screen(t)} + \hat{\beta}_{FP(1)}}}, \tag{5}$$

where  $\hat{\alpha}$  and  $\hat{\beta}$  are the maximum likelihood estimates from (1) and (2), respectively. For simplicity of exposition, suppose we want to estimate the cumulative risk of a subject who remains in the same age category  $i$  over  $n$  screenings with time between screenings in category  $j$ . (However the method can be extended to those who age from category  $i$  to category  $i + 1$ ). The estimated probability of at least one FP in  $n$  screenings (for  $n > 1$ ) is

$$p_{ij}(n) = 1 - (1 - a_{1i}) \prod_{t=2}^n (1 - a_{tij}), \tag{6}$$

The estimated survival time until the first FP in  $n$  screenings (for  $n > 1$ ) is  $1 - p_{ij}(n)$ .

To better quantify the cumulative burden of FP's we also estimate the expected number of FP's. The formula for the estimated expected number of FP's in  $n$  screenings varies with  $n$ . For example, for  $n = 4$  the estimated expected number of FP's is

$$e_{ij}(4) = pr(4FP's) 4 + pr(3 FP's) 3 + pr(2 FP's) 2 + pr(1 FP) 1, \tag{7}$$

where

$$pr(4 FP's) = a_{1i} q_{3|3ij}$$

$$pr(3 FP's) = a_{1i} q_{2|3ij} + (1 - a_{1i}) a_{2i} q_{2|2ij}$$

$$pr(2 FP's) = a_{1i} q_{1|3ij} + (1 - a_{1i}) a_{2i} q_{1|2ij} + (1 - a_{1i}) (1 - a_{2ij}) a_{3ij} q_{1|1ij}$$

$$pr(1 FP) = a_{1i} q_{0|3ij} + (1 - a_{1i}) a_{2i} q_{0|2ij} + (1 - a_{1i}) (1 - a_{2ij}) a_{3ij} q_{0|1ij} + (1 - a_{1i})(1 - a_{2ij}) (1 - a_{3ij}) a_{4ij}$$

and  $q_{h|fij}$  is the probability of  $h$  FP's over the last  $f$  time periods for a subject age  $i$  and with time interval  $j$  since the last screening. In this example

$$q_{3|3ij} = pr(3 FP's in last 3 screenings) = b_{2ij} b_{3ij} b_{4ij}$$

$$q_{2|3ij} = pr(2 FP's in last 3 screenings) = b_{2ij} b_{3ij} (1 - b_{4ij}) + b_{2ij} (1 - b_{3ij}) b_{4ij} + (1 - b_{2ij}) b_{3ij} b_{4ij}$$

$$q_{1|3ij} = pr(1 FP in last 3 screenings) = b_{2ij} (1 - b_{3ij}) (1 - b_{4ij}) + (1 - b_{2ij}) b_{3ij} (1 - b_{4ij}) + (1 - b_{2ij}) (1 - b_{3ij}) b_{4ij}$$

$$q_{0|3ij} = pr(0 FP's in last 3 screenings) = (1 - b_{2ij}) (1 - b_{3ij}) (1 - b_{4ij}),$$

$$q_{2|2ij} = pr(2 FP's in last 2 screenings) = b_{3ij} b_{4ij}$$

$$q_{1|2ij} = pr(1 FP in last 2 screenings) = b_{3ij} (1 - b_{4ij}) + (1 - b_{3ij}) b_{4ij}$$

$$q_{0|2ij} = pr(0 FP's in last 2 screenings) = (1 - b_{3ij}) (1 - b_{4ij}),$$

$$q_{1|1ij} = pr(1 FP in last screening) = b_{4ij}$$

$$q_{0|1ij} = pr(0 FP's in last screening) = 1 - b_{4ij}$$

An important special case occurs when the probabilities of FP do not vary with screening number. This case is important because it allows extrapolation to additional screenings. With  $a_{2ij} = b_{tij0}$  and  $b_{ij} = b_{ij1}$  for  $t = 2, 3, \dots, n$ , the estimated probability of at least one FP in  $n$  screenings is

$$p_{ij}(n) = 1 - (1 - a_{1i}) (1 - a_{2ij})^n, \tag{8}$$

and the estimated expected number of FP's in  $n$  screenings is

$$e_{ij}(n) = a_{1i} (1 + nb_{ij}) + \sum_{t=2}^{n-1} (1 - a_{1i}) (1 - a_{2ij})^{t-2} a_{2ij} (1 + (n-t)b_{ij}) + (1 - a_{1i}) (1 - a_{2ij})^{n-2} a_{ij}. \tag{9}$$

The most difficult part of implementation is computing the variance. The asymptotic variances are

$$var(p_{ij}) = \frac{\partial p_{ij}}{\partial \theta} var(\hat{\theta}) \frac{\partial p_{ij}}{\partial \theta} \quad \text{and} \quad var(e_{ij}) = \frac{\partial e_{ij}}{\partial \theta} var(\hat{\theta}) \frac{\partial e_{ij}}{\partial \theta}, \tag{8}$$

where  $\theta = (\alpha_0, \alpha_{age(i)}, \beta_0, \beta_{age(i)}, \beta_{time(j)}, \beta_{screen(t)}, \beta_{FP(k)})$ . By using computer software for symbolic derivatives [10], it is not hard to compute (8). Alternatively one could compute confidence intervals by using a bootstrap approach [11].

**Results**

We applied the methodology to data on 4 annual screenings in the Health Insurance Program of Greater New York (HIP) breast cancer screening study [13]. Starting in 1963, approximately 60,000 women were randomly assigned to either a study group invited for four annual mammograms and physical examinations or to a control group that received no screening within the study. Approximately 1/3 of the subjects in the study group refused the

**Table 1: Parameter estimates (standard errors) from logistic regression models for FP's**

	Narrow FP		Broad FP
	initial model	final model	final model
model for initial FP			
$\alpha_0$	-4.30 (.13)	-4.14 (.06)	-2.37 (.05)
$\alpha_{age}(2)$	.23 (.17)		.15 (.07)
$\alpha_{age}(3)$	.29 (.17)		-.01 (.08)
$\alpha_{age}(4)$	.13 (.19)		.05 (.08)
$\alpha_{age}(5)$	.00 (.22)		-.02 (.09)
model for subsequent FP's			
$\beta_0$	-4.81 (.18)	-4.89 (.06)	-2.85 (.07)
$\beta_{age}(2)$	.24 (.19)		-.02 (.07)
$\beta_{age}(3)$	.00 (.19)		-.16 (.08)
$\beta_{age}(4)$	-.32 (.21)		-.28 (.08)
$\beta_{age}(5)$	.07 (.21)		-.24 (.08)
$\beta_{time}(2)$	-.02 (.11)		-.13 (.05)
$\beta_{time}(3)$	.10 (.20)		-.22 (.09)
$\beta_{time}(4)$	.12 (.22)		-.08 (.09)
$\beta_{screen}(3)$	-.30 (.13)		-.11 (.05)
$\beta_{screen}(4)$	-.07 (.13)		-.22 (.06)
$\beta_{FP(1)}$	2.35 (.14)	2.34 (.14)	1.64 (.05)

first screening and received no screenings. Our analysis focused on the vast majority of screened women who were between ages 40 and 64. In HIP, our broad definition of a false positive was either (i) breast biopsy in which no cancer was detected or (ii) early re-examination with either clinical or radiological recommendation and no cancer detected on diagnostic work-up. Our narrow definition was (i).

Parameter estimates from the logistic regressions are presented in Table 1. The methodology answers two questions an individual might pose.

*Question 1.* "If I were to have  $n$  number of screening tests and stick to the schedule what are the chances that at least one will be a false positive?"

*Question 2.* How many unnecessary biopsies or work-ups am I likely to need if I start on the screening program and adhere to the schedule?"

Question 1 is answered by the estimates in (6) and (8). Question 2 is answered by estimates in (7) and (9). For an economic analysis Question 2 is particularly useful as it would help an analyst assign monetary costs to the cumulative burden of FP's. Both questions are clearly important to the patient.

Defining a FP as an unnecessary biopsy, we could not reject a model for (1) with only a constant (deviance = .27

on 4 d.f,  $p = .99$ ) nor a model for (2) with a constant and a parameter for previous FP's (deviance = 3.66 on 9 d.f.,  $p = .93$ ). Consequently we think it is reasonable to extrapolate beyond 4 screenings using (8) and (9). In answer to *Question 1*, the estimated probability of at least one FP in 10 screenings is .08 with 95% confidence interval of (.07, .09). In answer to *Question 2*, the estimated expected number of FP's in 10 screenings is .11 with 95% confidence interval of (.10, .12).

Defining an FP as an unnecessary work-up, most of the parameter estimates were statistically significant, so we kept all the covariates in the model. Because the parameter for screening number was included, we could not extrapolate beyond the 4 screenings that was the maximum number of screenings per subject in our data. For purposes of illustration we selected  $i = 1$  and  $j = 1$  for computing (6) and (7). In answer to *Question 1*, the estimated probability of at least one FP in 4 screenings is .21 with 95% confidence interval of (.20, .22). In answer to *Question 2*, the estimated expected number of FP's in 4 screenings is .34 with 95% confidence interval of (.32, .34).

As an ancillary investigation, we also fit a logistic regression for the probability of dropout as a function of age category, time interval since last screening, screening number, false positive on the last screening, false positive on an earlier screening, and interaction between of the two false positive variables. For the HIP data, when FP was defined broadly, there was no statistically significant asso-

ciation between FP history and dropout. For FP defined narrowly, there was a strong association between FP history and dropout.

### Discussion

Our methodology is applicable to any screening test recommended on a periodic basis for which data come from subjects with possibly different numbers of screenings. Ideally one would like data from a study in which subjects are representative of the general eligible population and clinicians are representative of the clinicians who would perform the screening in practice. Particularly when the FP is an unnecessary work-up, the clinicians may vary in the threshold used to determine a positive, as there is subjectivity to the interpretation of the test. When FP is an unnecessary biopsy, the variation among clinicians will likely be small because a high level of FP's is unacceptable [14]. In our data set there was no information on clinician. If there are data on clinicians, it should be incorporated into the analysis. If the number of clinicians is small, we suggest including a variable for clinician in the logistic regression. If the number of clinicians is larger, it is best to include a random variable for the effect of clinician. Unfortunately the simple logistic regression is not applicable and a more complicated model such as that in GW would be needed.

The assumption that dropout does not depend on future false positives could be violated if a subject drops out because of self examination results (so she goes to her regular physician) that would have led to future false positives. To avoid this violation of the assumption, one could ask women screened and women who dropped out if they found any lump on self examination. By including a covariate for lump on self examination, the dropout process depends on previous history and factors from the likelihood.

### Conclusion

We made three contributions. First we showed that previous methodology of GW did not require an unrealistic assumption about the dropout process. This makes the approach much more appealing. Second, we showed how to estimate the expected number of false positives, which we think is informative, in addition to the probability of at least one false positive. Third we presented a logistic regression formulation that is applicable for some data sets and is relatively simple to implement. Our approach can be applied to many types of cancer screening tests that are recommended on a periodic basis. It is useful for both advising individuals in a clinical setting and for health resources planning.

### Authors' Contribution

SGB wrote the initial draft and computed the results. DE created the appropriate tables of data from HIP and ran initial logistic regressions. BSK made substantial improvements to the manuscript.

### Competing Interests

None declared

### Additional material

#### Additional file 1

Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2288-3-11-s1.pdf>]

#### Additional file 2

Click here for file  
[<http://www.biomedcentral.com/content/supplementary/1471-2288-3-11-s2.pdf>]

### Acknowledgment

We thank Jian-Lun Xu, Victor Kipnis, and Philip Prorok for helpful comments.

### References

1. Baker SG, Kramer BS and Prorok PC: **Statistical issues in randomized trials of cancer screening** *BMC Medical Research Methodology* 2002, **2**:111 [<http://www.biomedcentral.com/1471-2288/2/111>].
2. Gelfand AE and Wang F: **Modeling the cumulative risk for a false positive under repeated screening events** *Statistics in Medicine* 2000, **19**:1865-79.
3. Elmore JG, Bargon MB, Mocerri VM, Polk S, Arena PJ and Fletcher SW: **Ten-year risk of false positive screening mammograms and clinical breast examinations** *New England Journal of Medicine* 1998, **338**:1089-1096.
4. Christiansen CL, Wang F, Barton MB, Kreuter W, Elmore JG, Gelfand AE and Fletcher SW: **Predicting the cumulative risk of false positive mammograms** *Journal of the National Cancer Institute* 2000, **92**:1657-1666.
5. McCann J, Stockton D and Godward S: **Impact of false-positive mammography on subsequent screening attendance and risk of cancer** *Breast Cancer Research* 2002, **4**:R11.
6. Kalbfleisch JD and Prentice RL: **The Statistical Analysis of Failure Time Data** Wiley: New York 1980.
7. Baker SG: **Regression analysis of grouped survival data with incomplete covariates: nonignorable missing-data and censoring mechanisms** *Biometrics* 1994, **50**:821-826.
8. Brown CC: **On the use of indicator variables for studying the time-dependence of parameters in a response time model** *Biometrics* 1975, **31**:863-872.
9. Efron B: **Logistic regression, survival analysis, and the Kaplan-Meier curve** *Journal of the American Statistical Association* 1988, **83**:414-425.
10. Wolfram S: *The Mathematica Book* 4th edition. Champaign: Wolfram Media and Cambridge: Cambridge University Press; 1999.
11. Efron B and Gong G: **A leisurely look at the bootstrap, the jack-knife, and cross-validation** *American Statistician* 1983, **37**:36-48.

12. Shapiro S, Venet W, Strax P and Venet L: **Periodic Screening for Breast Cancer, The Health Insurance Plan Project and Its Sequelae, 1963–1986** Baltimore, Johns Hopkins University Press 1988.
13. Baker SG: **The Central Role of Receiver Operating Characteristic (ROC) Curves in Evaluating Tests for the Early Detection of Cancer** *Journal of the National Cancer Institute* 2003, **95**:511-5.

### Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/3/11/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

