Original article

Vol. 1, 1 (May 2017)
ISSN: 2532-5876

DOI: 10.13133/2532-5876_6.8

Cancer biomarker discovery without assumptions about cancer biology: The double dip design

*Stuart G. Baker**

* *Biometry Research Group, National Cancer Institute, Bethesda, USA*

Corresponding author: Stuart G. Baker sb16i@nih.gov

Abstract

The biomarker pipeline to improve cancer screening begins with the discovery and validation of a cancer prediction model involving markers for the early detection of cancer in asymptomatic persons. Unfortunately, this biomarker pipeline has led to few markers for clinical use. An unappreciated reason for this lack of success is that standard discovery uses a convenience sample of specimens from persons with symptomatic cancer and no cancer. Standard discovery in a convenience sample implicitly makes a questionable assumption about cancer biology, namely, that highly predictive biomarkers in asymptomatic persons persist until symptomatic cancer arises when they outperform markers associated with symptomatic cancer. If cancer arises from a sequence of driver mutations and biomarkers are associated with driver mutations, this assumption may be plausible. However, if cancer arises primarily from changes in the microenvironment, the assumption is questionable. To circumvent the need for this assumption, I propose the double dip design. The double dip design starts with standard discovery in a convenience sample (as this is standard practice) followed by the usual validation sample of stored specimens from asymptomatic persons. If validation fails, it re-uses the original validation sample of stored specimens for more relevant biomarker discovery, followed by a second validation sample of stored specimens from asymptomatic persons. Recently developed statistical methods to reduce validation sample size make the double dip design feasible.

Keywords: biomarker, cancer screening, early detection, sample size, sensitivity, specificity, validation

Supplementary File: Validation Sample Size

Citation: Baker, SG, 2019, "Cancer biomarker discovery without assumptions about cancer biology: The double dip design", *Organisms. Journal of Biological Sciences*, vol. 3, no. 2, pp. 36-39. DOI: 10.13133/2532-5876_6.8

1. Introduction

There is a great need to develop new cancer screening modalities that decrease false positive screens, lessen overdiagnosis, and reduce cancer mortality. In general terms, a cancer screening modality is a cancer prediction model based on markers and risk factors. Markers are measurable indicators of biological state influenced by early-stage carcinogenesis. Examples include genomic markers in the blood, cyst fluid markers, antibody arrays, metabolites, protein markers in the urine, exosomes, circulating tumor cells, mutations in various genes, imaging results from mammography or ultrasound, and prostate specific antigen (Lippman et al, 2018, Sauter, 2017, Young et al. 2018). Markers can be collected at multiple times in each participant. Risk factors are measures of increased susceptibility to cancer, such as age, family history of cancer, and germline mutations.

The biomarker pipeline to develop a better cancer prediction model for use with screening has two phases: discovery and validation. In the discovery phase, investigators formulate a cancer prediction model, which involves both selecting (discovering) markers and fitting a model. In the validation phase, investigators use an independent sample to evaluate the performance of the cancer prediction model. A cancer prediction model is validated if it has good prediction performance (discussed more precisely later).

Discovery under the standard design involves specimens from persons with symptomatic cancer and controls without cancer. I call the discovery sample a convenience sample, because it is relatively easy for investigators to obtain specimens. Although the purpose of discovery in a convenience sample is to fit a cancer prediction model, the convenience sample provides no direct information for prediction.

Validation under the standard design involves stored specimens from asymptomatic persons. Investigators follow asymptomatic persons several years and measure markers in stored specimens from all participants

who developed cancer (cases) and a random sample of participants who did not develop cancer (controls) (Baker, Kramer, and Srivastava, 2002).

This standard design for biomarker discovery and validation has led to few clinical markers for early detection of cancer. Most markers for cancer early detection in widespread use were discovered between the mid-1960's and mid-1980's. These include carcinoembryonic antigen (CEA), prostate specific antigen (PSA), and carbohydrate antigen 125 (CA125).

Various researchers point to poor statistical methodology in study design as a likely explanation for the lack of success in finding and validating new biomarkers for the cancer early detection (Ransohoff, 2004, Pepe et al 2008, Ransohoff and Gourlay, 2010).

I discuss a more fundamental reason explaining the lack of success, namely that the standard design for cancer biomarker discovery requires a questionable assumption about cancer biology. In addition, I propose the double dip design which allows for cancer biomarker discovery with any assumptions about cancer biology.

2. Drawbacks of standard discovery

The standard discovery with a convenience sample implicitly assumes that highly predictive biomarkers in asymptomatic persons persist until symptomatic cancer arises when they outperform markers associated with symptomatic cancer. This assumption is consistent with the somatic mutation theory of cancer, that successive driver mutations lead to cancer, and the implication that biomarkers are associated with driver mutations. If the somatic mutation theory does not hold, standard biomarker discovery in a convenience sample would miss promising marker in asymptomatic persons in the following two ways.

First, a convenience sample would fail to discover a transient marker associated with preclinical cancer and not symptomatic cancer. Such a transient marker might signal the start of irreversible changes that lead to cancer. Possible examples of transient markers are markers related to stem cell signaling (Lipman et al, 2018) or intercellular signaling between stromal and epithelial tissue (Sonnenschein and Soto, 2016; Soto and Sonnenschein, 2011; Baker 2015, Baker 2018)

Second, a convenience sample would fail to discover a persistent biomarker of preclinical cancer that is masked by a better performing biomarker associated only with symptomatic cancer. For example, carcinoembryonic antigen (CEA) almost perfectly classifies the presence of colorectal cancer in a convenience sample involving specimens from persons with symptomatic colorectal cancer (Thomson et al, 1969). However, CEA poorly predicts the development of colorectal cancer in stored samples from asymptomatic persons (Thomas et al, 2015). Consider a biomarker M in asymptomatic persons that, unlike CEA, performs well for cancer prediction in asymptomatic persons. However, in a convenience sample, biomarker M does not perform as well CEA. A candidate cancer prediction model in the convenience sample that included both CEA and biomarker M would incorrectly indicate that biomarker M makes little, if any, contribution to cancer prediction in asymptomatic persons, simply because there is little room for improvement in cancer prediction with CEA in the model.

3. The double dip design

Baker (2009) proposed a discovery phase using stored specimens from asymptomatic persons. Although this design would avoid the assumptions with standard discovery in a convenience sample, it is unacceptable to most investigators. Many investigators are reluctant to perform discovery using stored specimens from asymptomatic persons, thinking it wastes precious specimens (ignoring the downside of wasting stored specimens to validate unpredictive markers discovered in a convenience sample).

The double dip design circumvents the limitations of the standard design in a practical manner (Figure 1). The double dip design starts like a standard design with discovery in a convenience sample and validation using stored specimens from asymptomatic persons.

The key to the double dip design is the next step. If the validation sample indicates poor performance of the prediction model formulated in the convenience sample, the double dip design re-uses the prospective validation sample as a second-chance discovery sample -- a procedure which I call the double dip. The double dip yields a cancer prediction model based on markers in stored specimens from asymptomatic persons –which is what is needed for relevance to early detection. To evaluate the second-chance cancer prediction model, the double dip design requires a second prospective validation sample using stored specimens from asymptomatic persons.

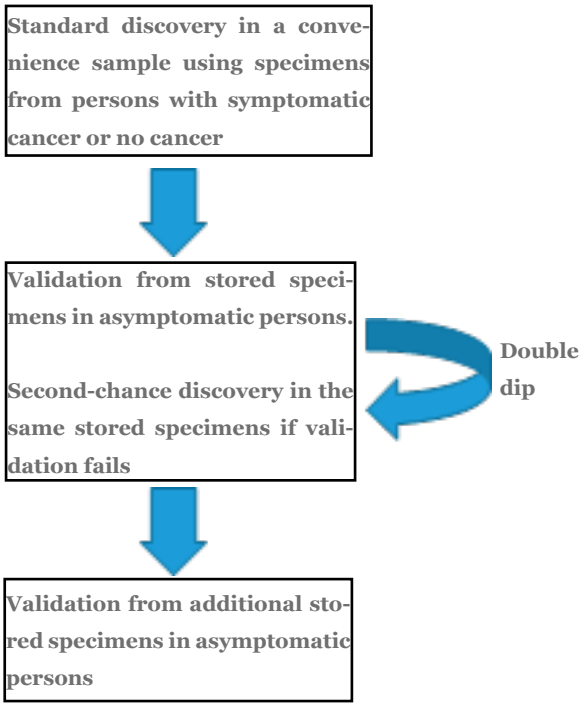

Figure 1. Double dip design

4. Sample size

The main drawback to the double dip design is the need for two validation samples of stored specimens. Fortunately, a recently developed statistical method yields reasonable sample sizes for validation (Baker, 2019). The key is to estimate sensitivity (probability of a positive cancer prediction given develop cancer) imprecisely and to target 100% specificity (probability of a negative cancer prediction given no cancer arises in the study). The high specificity ensures a high positive predictive value (probability cancer arises in the study given a positive prediction) regardless of the sensitivity.

Under Scenario 1, the target values are 80% sensitivity with lower bound of 50% and 100% specificity with lower bound of 99.5%. Under Scenario 2 (which is easier to achieve), the target values are 50% sensitivity with lower bound of 20% and 100% specificity with lower bound of 99.5%. Both scenarios require specimens from 12 cases (persons who develop cancer) and 740 controls (persons who did not develop cancer). Under Scenario 1, the cancer prediction model is validated (achieving target performance) if at least 9 out of 12 case specimens are positive when 0 control specimens are positive. Under Scenario 2, validation requires at least 5 out of 12 case specimens to be positive when 0 control specimens are positive. See the online supplementary appendix for sample size calculations. Table 1 (which applies to both scenarios) shows the sample sizes for the total number of persons contributing specimens in each validation sample.

Probability of developing cancer during the study	Validation sample size
1.0%	2000
1.5%	1300
2.0%	1000
2.5%	800

Table 1. Validation sample sizes. All designs are based on 12 cases and 740 controls to yield target performance.

5. Discussion

A potential limitation of the double dip design is that the second-chance discovery sample (which is the original, re-used, validation sample) may involve too few cases for adequate discovery. If this is a concern, investigators could double its size and still have a reasonable sample size in many scenarios.

An important determinant of sample size is the probability of developing symptomatic cancer in the study. As shown in Table 1, higher probabilities of developing symptomatic cancer in the study translate into smaller sample sizes. Therefore, investigators should collect specimens from populations at high risk of developing symptomatic cancer with the understanding that results strictly only apply to high risk persons.

The double dip design can increase the efficient the use of stored specimens in trials where biomarker discovery and validation are not the main goals. Investigators collected stored specimens in two large prevention trials with lung cancer incidence as the primary endpoint, the Alpha-Tocopherol Beta Carotene Lung Cancer Prevention Trial (ATBC) (ATBC Cancer Prevention Study Group, 1994) and the Beta-Carotene and Retinol Efficacy Trial (CARET). (Omen et al 1996). In using these stored specimens to predict prostate cancer, Baker (2000) essentially performed a double dip design with discovery in ATBC stored specimens and validation in CARET stored specimens.

The formulation of the cancer prediction model can involve the “discovery” of markers from high-dimensional data such as might arise from microarrays or other -omics approaches. It can also involve reverse time models to better accommodate varying numbers

of markers collected over time in each person (Baker and Tockman, 2002).

There is a growing appreciation of the advantage to using stored specimens for discovery. Ransohoff (2017) wrote, "As stated by one observer 'We need to turn conventional wisdom on its head' and use precious specimens far earlier than we currently do (Z Feng personal communication)." Until now there has been no acceptable path to the ideal discovery using stored specimens. The double dip design provides such a path.

Acknowledgments

This research was supported by the National Institutes of Health. The statements contained herein are solely those of the author and do not represent or imply concurrence or endorsement by the NIH.

References

ATBC Cancer Prevention Study Group, 1994, The Alpha-Tocopherol, Beta-Carotene Lung Cancer Prevention Study: Design, methods, participant characteristics, and compliance. *Annals of Epidemiology,* vol. 41, pag. 1-10.

Baker SG, 2000. Identifying combinations of cancer biomarkers for further study as triggers of early intervention. *Biometrics,* vol. 56, pag. 1082-1087.

Baker SG, 2009, Improving the marker pipeline to develop and evaluate cancer screening tests. *Journal of the National Cancer Institute,* vol. 101, pag. 1116-1119.

Baker SG, 2015, A cancer theory kerfuffle can lead to new lines of research. *Journal of the National Cancer Institute,* vol. 107: dju405.

Baker SG, 2018, The detached pericyte hypothesis: A novel explanation for many puzzling aspects of tumorigenesis. *Organisms. Journal of Biological Sciences,* vol. 2, pag. 25-41

Baker SG, 2019, Cancer screening markers: a simple strategy to substantially reduce the sample size for validation. *Medical Decision Making*, vol. 39, pag. 130-136.

Baker SG, Tockman MS, 2002 Evaluating serial observations of precancerous lesions for further study as a trigger for early intervention. *Statistics in Medicine*, vol. 21, pag. 2383-2390.

Baker SG, Kramer BS, Srivastava S, 2002, Markers for early detection of cancer: Statistical issues for nested case-control studies. *BMC Medical Research Methodology,* vol 2: 4.

Diamandis E P, 2012, The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Medicine* vol 10: 87.

Lippman SM, et al, 2018, AACR White Paper: Shaping the Future of Cancer Prevention - A Roadmap for Advancing Science and Public Health. *Cancer Prevention Research (Phila)*, vol. 11, pag. 735-778.

Omenn GS, et al, 1996, Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Retinol Efficacy Trial. *Journal of the National Cancer Institute*, vol. 109: djx060.

Pepe MS, et al, 2008, Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *Journal of the National Cancer Institute,* vol 100, pag. 1432–1438.

Ransohoff DF, 2004, Rules of evidence for cancer molecular-marker discovery and validation. *Nature Reviews Cancer*, 4, 309-314.

Ransohoff DF, 2017, Improving the clinical validity of biomarker research in cancer detection, in S. Srivastava, ed. *Biomarkers in Cancer Screening and Early Detection*. West Sussex U.K: Wiley, 209-219.

Ransohoff DF, Gourlay ML, 2010, Sources of bias in specimens for research about molecular markers for cancer. *Journal of Clinical Oncology*, 28, 698-704.

Sauter ER, 2017, Reliable biomarkers to identify new and recurrent cancer. *European Journal of Breast Health*, 13, 162–167.

Sonnenschein C, Soto AM, 2016, Carcinogenesis explained within the context of a theory of organisms, *Prog Biophys Mol Biol*, vol. 122, pag. 70-76.

Thomas DS, Fourkala EO, Apostolidou S, et al, 2015, Evaluation of serum CEA, CYFRA21-1 and CA125 for the early detection of colorectal cancer using longitudinal preclinical samples. *British Journal of Cancer*, vol. 113, pag. 268–274.

Thomson DM, KrupeyJ, Freedman SO, Gold P, 1969,The radioimmunoassay of circulating carcinoembryonic antigen of the human digestive system. *Proc. Natl Acad. Sci. USA*, vol 64, pag.161–167.

Young MR, et al. (2018) Validation of markers for early detection of pancreatic cancer summary of The Alliance of Pancreatic Cancer Consortia for Markers for Early Detection Workshop. *Pancreas*, vol 47, pag. 135-141.

Soto AM, Sonnenschein C, 2011, The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory, *Bioessays*, vol. 33, pag. 332-340.