Minding Your P's and Q's

TWO NOT-QUITE-JURASSIC CLASSICS, AND SOME MATHEMATICA PROSTATICA

Michael W. Kita, MD

This issue of the Journal contains three abstracts (two cardiac, one cancer) and a report on PSA (Prostate-Specific Antigen). Each of these articles has some interesting mathematical features deserving of comment.

Classics

Singer's two abstracts are classics, in several senses of the word. They are the first two abstracts to carry the L-code, which was set aside for important older studies not previously abstracted. They also have unusually extensive data with unusual completeness of Follow Up. The original articles on which the abstracts are based were published in the 1950's, not long after Berkson and Gage's key paper, and they represent two early examples of the application of actuarial life-table methods to longitudinal studies by clinicians.

Abstract 350L1 is a unizonal multiserial multisecular study of angina patients. It is a historical-prospective (“retroactive”) study by design: its 6882 angina patients were assembled from record-review of Mayo Clinic files, and then followed forward in time. It is unizonal, since all subjects were from a common and geographically discrete source. It is multiserial, or cross-sectional, because some of the anginal patients were newly diagnosed while others had been diagnosed with angina for 10 years and longer. It is also multiserial in nature since its 17-year wide rolling-entry period has the effect of creating multiple small series from the succession of waves of slightly dysynchronous (non-contemporaneous) enrollments. It is multisecular in view of both the long enrollment period and the long duration of the study (23 years).

Abstract 350L2 is a unizonal, uniserial, multisecular study of angina patients. It is a prospective study, ambitiously begun in 1920 by a single investigator who assembled this series of 456 angina patients from his own practice over a span of 11 years, and then methodically followed them all for another 25 years. This makes Paul Dudley White's study not only classic, but almost heroic in effort. Only five patients were lost to FU (none sooner than 14 years, and the last one after more than 26 years of Follow Up) and a mere six patients were still alive at the end of the study - showing how close White came to conducting his own personal 36-year single-decrement mortality study. Of those patients who were followed to completion of study, 25 years FU was the minimum duration of FU, with 36 years being the maximum. This kind of FU is heroic by modern standards, and while it is still possible to conduct such a unizonal, uniserial study via a single center like Framingham, it is not likely to be duplicated by any single investigator again.

Abstract 158M1 (the pancreatic cancer study) may not be as classic as the two angina studies in date, but what it lacks in antiquity it makes up for in ubiquity. It is multizonal, both by virtue of being a multicenter study, and by being a multinational one at that. Like the Mayo Clinic study, it is retroactive in design, multiserial and multisecular. It identified its subjects by record-review, then followed them prospectively through subsequent records, until end point (target event), end of study, or loss to FU. It is multiserial both in respect to the wide enrollment period (43 years) and to the variability in time-since-diagnosis with chronic pancreatitis (CP). (All subjects had a common zero-time of medical center diagnosis with CP, but varied in how long CP may have been present prior to presentation to those medical centers). It is multisecular in view of both the long entry-period and the long study-period.

Besides differences in study design (see Table 1), each of the three abstracts offered unique methodologic idiosyncrasies to challenge the abstractor.

Discarded Data

The Mayo Clinic angina study and the pancreatic cancer study each chose to discard portions of their exposure data in analyzing and reporting their findings. In the Mayo Clinic Study, only patients traced to the end of a full study-interval (5 years, 10 years, etc.) were included in exposure; those lost to FU or withdrawn alive prior to completing interval-FU were excluded from analysis. A study which excludes partial contribution to exposure in this fashion is called an ad hoc study. Ad hoc mortality rates may overestimate interval mortality, compared to aggregate rates for the same interval. In Abstract 350L1, Singer reconstructed the missing expo-
sured from w (withdrawal) information reported in the source article. He added back to his Table 350L1-2 the 3245 person-years that were originally discarded, and arrived at true aggregate q’s for the first fifteen annual durations.

In the pancreatic cancer study, the authors discarded the first two years of exposure, artificially truncating the exposure in the interests of getting a clean base-line separation between CP and pancreatic cancer. They did not want to include as subjects those persons whose CP might have been due to already existing pancreatic cancer. By instituting a two-year "elimination period" on countable cases of pancreatic cancer, they attempted to bolster their case-definition of chronic pancreatitis so that they could conclude that any cases of pancreatic cancer arising in their CP subjects would have the cause-and-effect arrow running from CP to pancreatic cancer. At the same time, however, they chose to discard the first 2-years worth of accruing exposure as well. The effect of discarding exposure was to artificially increase the observed morbidity rates. Fortunately, the missing exposure could be calculated, and the corrected aggregate rates reconstructed (Table 158M1-2).

In both abstracts 350L1 and 158M1, the effect of restoring the discarded exposure was to reduce the magnitude of the comparative mortality and morbidity differences. In the angina study (350L1) the omitted exposures are spread over 20 durations, spreading and diluting the impact of the undercounted exposure. In the pancreatic cancer study (158M1), the omitted exposure is a block of exposure, up front, and its omission reduces the magnitude of morbidity ratios for each of the subgroups in any study period commencing with the true time-zero of the study.

Other Observations

Abstract 350L1 was also interesting for its advancement of q’ by duration. After deriving interval q’ from annual cumulative P’ information provided, Singer estimated that an approximately 8.5% annual rate of increase was being applied, in order to yield the column of q’ values for durations 1-15 in Table 350L1-2. Such a rate of increase is close to that seen in general population tables, and would arise from the beginner’s assumption that survivor mean-age increases one full year for each year of elapsed time. Empirically, this is seldom the case. In clinical studies, older members of a cohort decease faster than younger members, and the mean-age of the survivorship as a whole ages at something less than a full year for each year of elapsed time. In general, advancement of q’ of 6% or less per year is a better estimate, unless sufficient data is available to permit recalculation of mean q’ year-by-year, according to age/sex rematching. The effect of the 8.5% assumption by the study authors is to tend towards overestimation of q’ and d’, especially at the longer durations, with resultant underestimation of MR and EDR.

Singer also notes that this 8.5% compounding of the expected mortality rate makes the denominator of MR double every eight years, causing the MR to continue to shrink even when significant annual EDR’s are still being experienced. For instance, while the EDR continues to average 40-50 deaths per 1000 between durations 6-12, the MR steadily shrinks from 245% to 190%. And for durations 8-10, when the EDR is constant at the non-trivial rate of 43 per 1000, the MR falls from 205% to 190%. While MRs may be fairly meaningful over short study durations (e.g. 5-10 years) or in studies of younger-age groups where q’ is small and relatively flat (when measured per 1000), they can be somewhat misleading at higher-ages or over extended (e.g. 20-year) study-frames. The EDR may be the more meaningful statistic in these latter circumstances. Fuller consideration of these points can be found in two earlier papers by Singer.4,5

Just When You Think You’ve Seen Everything

While the Paul Dudley White article did not discard any data, it did choose to report its rates and ratios by rather unique FU durations – the customary five-year intervals plus a half-year. The extra half-year was rationalized by the authors as necessary to adjust for the fact that age to nearest birthday and age at last birthday aren’t always the same. For instance, in a group of 40-year-olds, some may be nearer age 40 (last birthday) and some nearer 41 (next birthday) and thus subject to different at-age-x mortality rates. This actuarial distinction, while sometimes applied to persons or to insurance policies, is unique in being applied here to durations of FU. Accordingly, Singer used a conversion-factor to "normalize" reported rates and ratios to customary quinquennial periods, and then showed the data in standard format.

Tumor Markers

Also appearing in this issue is a distillation by Chambers of much of the current thinking about prostate cancer risk and the potential of PSA (Prostate-Specific Antigen) as an evaluative tool in the assessment of this risk. The usefulness and the limitations of PSA exemplify many of the things that currently make tumor markers a much-debated subject.

By way of background, the reader will find much useful information in the Proceedings of the Academy’s Tumor
Marker Forum, published last year as a JIM supplement, and in Pokorski's excellent discussion of tumor marker issues from an insurance perspective, published in the prior issue of this Journal. Bergstrom's article on the protective value of PSA testing contains a still-useful cost/benefit analysis from a life insurance perspective, and Mills has provided a framework for a more general protective value model encompassing morbidity as well as mortality expectations.

In the U.K., malignant neoplasms account for about 5% of hospital occupancy, 2.5% of GP consultations, and 20% of lost years of life expectancy. Malignant disease has similar social and fiscal implications in other countries. The potential value of having an effective means of detecting those cancers for which early intervention can be life saving and cost sparing could be substantial. Will tumor markers fulfill this promise?

Tumor markers are tumor-cell constituents that are assayable in tissue or body fluids. The constituents may be cell products (possessing biologic activity, such as enzymes or hormones) or cell components (antigens, receptors, structural proteins, or other such markers). They may be tissue-bound, but amenable to immunohistochemical detection. Or they may be circulating substances, with certain clearance characteristics, amenable to bioassays of various types. Their detection may be qualitative (present or absent) or quantitative (measurable amounts). Their clinical value will depend on how unique their association is, both with a malignant process (tumor-specific) and with a particular site of origin (tissue or organ-specific). There needs to be something sufficiently differential about the tumor marker—something unusual about the type of molecule it is, or the amounts in which it is being made—that distinguish it from non-malignant states (growth, repair, pregnancy, hyperplasia, and biological variation). Since all cells contain the same original genetic endowment, cells of many types may be capable of expressing the "protein" we wish to construe as a tumor marker—facultatively, if not constitutively. (Somatic mutations or viral oncogenes may target particular cell lines and confer organ specificity.) Ultimately, the "tumor marker" is one of the protein (and sometimes protein) expressions of certain oncogenes residing in that person's genome. Or more accurately, it is the final dysregulatory expression of the dynamic interplay of oncogenes and tumor suppressor genes—in the case of prostate cancer, perhaps ones lying near the PSA site on chromosome 19. The ideal tumor marker is one that is expressed uniquely by cancer cells, and can be detected with reliability at the time that the tumor achieves biologic importance.

Prostatic cancer can be an indolent condition—a slow-growing malignancy occurring at elder ages—and a tumor marker intended to detect prostate cancer would ideally be one that identified early malignancies that would lethally progress but for an intervention which is curative at that early stage. In other words, it would not merely mark the presence of advanced or incurable disease, nor would it mark the presence of inconsiderential cancers that would never progress to life-threatening disease in the subject's lifetime.

Prostate Cancer

Chambers summarizes the growing health significance of prostate cancer in the U.S., where the demographics of aging and secular improvements in average life expectancy (reduction in the magnitude of some of the competing causes of mortality) have permitted men to live long enough to develop and manifest prostate cancer. And age remains one of the chief "risk factors" for this malignancy (see Table 2). But it is a growing world health problem, as recent data from the World Health Organization shows (see Table 3). Secular trends have been towards both increased incidence rates and increased mortality rates.

It is said that prostate cancer is generally so slow-growing, and slowly-progressing, that a person is more likely to die with it, than to die from it. And while it is true that prostate cancer lacks the striking case-fatality rates of something like pancreatic cancer, it is nonetheless the case that disease-specific mortality rates are still quite high. Risk of cancer death is 25-30% at 20 years from first diagnosis for localized disease, and 75-85% for node-positive disease. (Approximately 30% of patients who appear to have localized disease when initially detected will in fact have pelvic lymph node metastases when staged.) Even localized prostate cancer has rather high residual EDR's at durations beyond 10 years since diagnosis, with EDR ranging from 20 to 50/1000/year.

Sardino has estimated that for a 50 year-old male in the US, the risk of developing clinically evident prostate cancer is about 3/1000 per year, and the risk of dying of prostate cancer, about 0.9/1000 per year. This yields a lifetime risk of developing a clinically detected prostate cancer of 6-8%, and a lifetime risk of dying of prostate cancer of 2.89%.

PSA

Prostate Specific Antigen is a 34 kilodalton glycoprotein that appears to originate exclusively from prostate tissue. It was first isolated from seminal fluid in 1971 and
was then called gamma-semiprotein. When its specificity for prostate-tissue was established, it was renamed PSA. PSA is secreted into the lumen of the prostate ducts and is present in seminal fluid in high concentrations (1-5mg/ml). It is a kallikrein-like enzyme with serine-protease activity, the biological function of which appears to be the liquefaction of the seminal coagulum formed during ejaculation, thereby facilitating the liberation of spermatozoa. Its 240 amino-acid peptide chain exists in several isomeric forms, and is coded for by a gene on chromosome 19 that has been fully sequenced.

By virtue of its low molecular weight, PSA can diffuse into the circulation where it is cleared with a half-life of 2 to 3 days. Levels in the circulation are typically less than 4 ng/ml (by monoclonal immunoassay) and reflect the volume of secretory prostatic tissue as well as permeability and hydrostatic factors (e.g. distortions of duct architecture from crowding, fibrotic changes, etc). PSA levels are also subject to androgenic hormonal influence (both testosterone levels and adrenal androgens, as well as exogenous anabolic steroids.) Prostatic massage can cause transient elevations of the circulating PSA levels, and prostatic injury (trauma, biopsy) can cause 50-fold elevations for as much as two weeks.15 Digital rectal exam (DRE) is not supposed to elevate PSA levels when the baseline level is itself normal (eg, <4 ng/ml) but may further elevate already elevated levels (especially when the baseline level was already greater than 10 ng/ml).16 However, some controversy exists as to whether delayed elevations can occur as a result of DRE, even if immediate elevations do not. Diurnal variation is generally considered small, but day-to-day variation up to 10% has been reported by some. The current monoclonal PSA test generally has good accuracy and precision, with standard errors of only ±4-5%.

PSA is secreted by normal prostate tissue, hyperplastic prostatic tissue, and prostate cancer cells. Some poorly differentiated prostate cancers may hypossecrete PSA, and result in circulating PSA levels that appear "normal." But most prostatic cancers secrete generous amounts of PSA – more than 10 times, gram for gram of prostate tissue, what normal prostate cells secrete.15 Hence, in theory, a small volume of tumor cells should begin to differentiate itself from the baseline secretory output of the surrounding gland at a certain critical mass. Unfortunately, that surrounding gland can itself have a volume of 20 to 40 cc or more (sonographic volume). The size of the "normal" prostate gland increases somewhat with age, as does the likelihood of developing some degree of BPH. So PSA determinations must be interpreted against a background of considerable "noise."

Nonetheless, the American Cancer Society (ACS) issued a guideline in November 1992 recommending annual DRE beginning at age 40 for prostate screening, and annual PSA at ages 50 and up. It recognized that abnormal elevations (>4ng/ml, monoclonal) could be associated with BPH or prostatitis in the absence of cancer, and advised transrectal ultrasound (TRUS) for further evaluation of abnormal DRE and/or abnormal PSA. A negative TRUS and/or biopsy would return the person to physician surveillance. The ACS also suggested earlier screening among blacks and those with positive family history, and generally advised that screens ought to be those with a 10-year or greater life expectancy, in order to receive a presumed benefit from screening. They ended on the cautionary note that reduction in mortality from screening has yet to be documented.

Catalona's Study

In one of the most widely cited studies of the value of the PSA in screening for prostate cancer, Catalona,37 et al., looked at two groups of men 50 years and older. 1653 ostensibly healthy "invited volunteers" (mean age 67) formed their "study group" and a series of 300 consecutive prostate-biopsy patients comprised their "comparison group" (mean age 68.5). (The authors chose, opposite to the usual convention, to call their "symptomatic" group the comparison group, and their "healthy" group the study group, which can be a little confusing to the reader). 65 members of the biopsy group had no PSA determination done, and were excluded from study. Of the remaining 235 – those who had both PSA and biopsy – cancer was found in 61 (prevalence 26%). Thus, in this group, PSA had a sensitivity of 79%, a specificity of 59%, and a positive predictive value of 40%. But 13 of the cancers (21% of the 61 cancers found in the group) occurred in individuals with PSA <4.0 ng/ml.

137 men (8%) of the "study group" had PSA's >4.0 (2% of them had PSA greater than 10), but of these men, only 112 submitted to biopsy. Of those biopsied, 37 were found with cancer. The 19 men with PSA between 4 and 10 ng/ml were found to have clinically localized cancers, but 7 of them were reclassified as advanced disease when they were surgically staged. None of the study group members with PSA <4 were biopsied, so exact information on "false negative" PSA is not available on this group. But assuming a comparable sensitivity and specificity to that found in the comparison group, and using the prevalence found for the study group (2.2%)
the positive predictive value of PSA >4 for prostate cancer for the study group would only be 4%.

Neither group is representative of a general population. The comparison group had a variety of symptoms and findings suspicious enough for cancer to warrant biopsy. The study group consisted of men responding to a press release soliciting volunteers for prostate cancer screening, and even supposing that the selection bias was small, individuals with a history of any prostatitis were excluded. Moreover, while the title of the article was "Measurement of PSA in serum as a screening test for prostate cancer," the sensitivity and specificity values for PSA in the detection of cancer were calculated on the non-screening group, (who represented a consecutive series of prostate-biopsy patients). And the members of the "study group" undergoing biopsy weren’t just PSA-tested, but serially tested and DRE-TRUS screened before undergoing any biopsy. The true "cancer status" of all other subjects in the "screened" group is unknown. Thus, the findings, while interesting, have limited generalizability, and demonstrate the "blur" that exists between PSA results that are indicative of malignancies, and those that are not.

Gland Math

A tumor volume of .5 cc corresponds roughly to a transverse diameter of approximately 1 cm if the tumor is generally spherical. \(\frac{4}{3}\pi r^3\). A tumor this size will result from approximately 30 doublings of an original cancer cell. A tumor mass 1 cm in diameter will thus contain roughly a billion cells. When tumor volumes are determined by ultrasound, they typically are estimated from a formula like \((l \times w \times h \times 0.52^3)\) using maximum l, w, and h values from longitudinal and transaxial views. (A spherical object of diameter d typically has a volume about half that of a rectangular object of dimensions d x d x d, hence the factor of 0.52).

False negatives

All three methods of primary evaluation for possible prostate cancer (PSA, DRE, and TRUS) are subject to false-negative results. For PSA, results in the normal range can occur if the secretory tumor volume is too small to produce an elevated result, or if the tumor is too poorly differentiated to have significant secretory capacity. It is estimated that as many as 32% of early stage tumors may lack the critical volume necessary to produce abnormal elevations of the PSA, and that 3% or more of tumors may be so poorly differentiated as to have PSA levels hidden in the "normal range." DRE can be falsely negative if the tumor nodule is anterior, or not projecting dorsally or laterally in a manner that can be discerned by palpation, or is smaller than about 7 to 8 mm in diameter (approximately .4 cc in volume). Fortunately, most prostate cancers arise in the peripheral zone of the prostate gland, and a high percentage of them will eventually protrude dorsally or laterally. TRUS can be falsely negative if the tumor is isoechoic, or insufficiently hypoechoic to be differentiated from adjacent tissue. False negatives diminish the sensitivity of a test, and the various sources of false negatives described here – which are essentially irreducible – account for the inability to generate test sensitivities approaching 100%.

False Positives

False positives are generally less of a problem for DRE and TRUS than for PSA. A palpable nodule on DRE may turn out to be an anatomic alteration associated with non-malignant disease or prior prostate injury/damage. A skilled examiner may well be able to discern such differences. In fact, for DRE, the "screening result" is not simply "nodule, present or absent", but a more complex assessment involving clinical judgment about shape, symmetry, texture, induration, nodularity, and other impressions beyond simply presence or size. For most tests, the overlap of diseased and disease-free populations typically means that improved specificity can only be gained at the expense of reduced sensitivity (see "Drawing Conclusions from Test Results"). But DRE is an example of a test for which both specificity and sensitivity can go up, in the hands (actually, the fingertips) of the highly skilled examiner. TRUS can result in false positives despite the fact that those reading and interpreting the images have a high initial skill level already. The prostate gland has a complex architecture, and vascular structures, cysts, and artifacts can all have a hypoechoic appearance. False positives for the PSA (when used as a threshold test or single cut-off test) are due to the considerable overlap between malignant and non-malignant prostatic disease.

Cost-Benefit Analysis

A cost-benefit model for assessing the value of screening has been developed\(^{19}\) and modified for prostate cancer screening.\(^{20}\) The role of false positives and false negatives can be seen from looking at the cost benefit formula, which has the following five terms:

\[
V = pmwB + (1-p)cR-S-[pm+(1-p)(1-c)]w(T+A)-(1-p)(1-c)F
\]

Where

\[
\begin{align*}
(1) & \quad \text{V} = \text{Expected Value of a Screening Test} \\
(2) & \quad \text{pm} = \text{Probability that a Cancer is Present} \\
(3) & \quad \text{w} = \text{Weight of a Cancer Diagnosis} \\
(4) & \quad \text{T} = \text{Cost of a True Positive} \\
(5) & \quad \text{F} = \text{False Positive Rate}
\end{align*}
\]

The capital letters are values of costs or benefits, and the lower-case variables are probabilities. Table 4 explains how these variables are defined. The first two terms of
the formula are "benefits" and the remaining three subtracted terms are "costs."

The first term is the benefit of treatment. It consists of the future medical savings from treating earlier cancer, plus the savings in wages-not-lost, plus the reduced suffering from prevention of advanced disease. The value of B is multiplied by p(nw), the probability that the person has the disease, that the test is positive, and that the person returns for treatment.

The second term is the benefit of reassurance, which is figured as the probability of a true negative result times the money expended to have the screening test done. The third term is the screening cost, which is the cost S of the test, incurred 100% of the time for each person screened. The fourth term is the adverse reaction cost, which is the cost of unnecessary treatment and the cost of any complications from invasive tests or toxic therapies that result from both true positives and false positives. The last term is the false positive cost, the psychic toll or suffering which occurs with a probability equal to that of a false positive result.

Table 4 also shows the probability and dollar-value assumptions (or in some cases, the range of assumptions) used by the ACS National Prostate Cancer Detection Project. Some of its costs are imputed costs, and other are time-dependent averages. Moreover, costs of treatment and savings for early-detection will depend on the relative efficacy of invasive treatments as compared to "watchful waiting" for early stage prostate cancer, as a recent decision analytic model has suggested.

Because of difficulties in pricing the less tangible factors, the researchers for the ACS Prostate Project made the simplifying assumption of setting R and S equal to the costs of the detection tests in question, and setting F*A, the cost of adverse treatment effects and unnecessary therapy. Other assumptions - based on utility values derived from patient preferences and perceived value - are possible.

A population prevalence of 5.3% was also assumed, but the researchers acknowledge that the prevalence of screen-detectable prostate cancer is unknown.

Using the cost benefit formula above, and the valuation assumptions of the prostate project researchers, PSA screening for prostate cancer failed to show net benefit for PSA at levels >4 if the threshold for abnormal is placed at 4 ng/mL. For clinical decision-making the "reliability" of the PSA test is coupled not just to its ability to detect prostate cancer, but its ability to detect early prostate cancer, when disease outcome and burden can be most greatly impacted.

However, the ACS concluded that if a more specific PSA test (one with higher specificity, and fewer false positives) could be developed, the combination of PSA and DRE could represent an ethical and economical detection choice for individual patients in consultation with their physicians.

For other types of risk-decisions, such as insurance decisions based on excess morbidity or mortality compared to standard expectations, a PSA level associated with "excess risk" may or may not be identical to the level chosen by clinicians or public health experts for clinical decision-making. Moreover, rising overall rates of prostate cancer prevalence bear on such determination.

Protective Value

Protective value calculations are another kind of cost-benefit analysis. Instead of looking at screening as costs incurred versus costs-saved/benefits conferred from the standpoint of clinical decision-making, protective value looks at mortality and morbidity cost savings from an insurance perspective. It then asks, in light of certain interest rate, lapse rate, and other assumptions, whether the costs prevented (the savings attributable to instituting the test) justify the costs of testing. Not part of the protective value calculation itself, but part of any decision to embark on new or additional insurance testing, are a host of technical, regulatory, ethical, and market implications, as noted by Chambers, Pokorski, and others.

Combinatorics

Tests can be obtained for a variety of purposes, in various populations, alone or in combination with other tests, performed serially or simultaneously. In each such scenario the results will be different in terms of net information content. Table 5 shows some of the ways these can be conceptualized for PSA. PSA will have paired values of sensitivity and specificity for prostate cancer that will depend on the gold standard used (e.g. presence of any cancer on biopsy versus presence of localized cancer) and the details of the selection of the "with-disease" and "without-disease" populations (comparability with regard to age, DRE status, biopsy technique, histologic review, etc.). And the sensitivity-specificity pair will take on different values depending, for a continuous variable like PSA concentration, on the cutoff level chosen to define a "positive" or suspicious
result, and the degree of overlap of the with-disease and without-disease populations.

When tests like PSA are used as stand-alone tests, their information content or predictive value for particular diseases will depend on the pre-test likelihood of disease in the subject. If the only information available is age, then some general population prevalence may be assumed. If specific risk factors are known, or the results of a recent physical and DRE, then a different pre-test likelihood may pertain. If the results of PSA are used in conjunction with other test results, like DRE or TRUS, the combinatorics of the separate sensitivities and specificities for each test will depend on whether a decision is to be based on either test being positive, or on both tests being abnormal (Table 6).

Refining the Tools

PSA has many problems associated with its use – alone as a diagnostic or screening test. But it does not have to be interpreted as an isolated result, referenced only to an age-based prevalence assumption. The threshold value for defining a meaningful abnormal could be placed at some level other than 4 or 10, with consequent repartitioning of false positives and false negatives. The PSA can be standardized to prostate gland volume, as the PSAD attempts to do. (The same TRUS that would provide PSAD information would presumably also furnish information as to isoechocility.) The PSA level could be interpreted serially, referenced to the rate of change or to some absolute amount of interval increase. The PSA result could be interpreted in concert with DRE and/or TRUS findings (in either/or fashion when discordant, or according to some other combinatorial rules.)

In the final analysis, a number of questions need to be considered. Is the test a screening test – intended to raise an index of suspicion and move the individual into a further evaluative process – with final diagnostic judgment suspended until the evaluative/diagnostic process has been completed? Or is the test imagined to be a diagnostic test, capable by itself, or in conjunction only with other baseline information, of yielding a positive predictive value high enough to be considered “diagnostic”?

As to the role of PSA in prostate cancer, one can ask:

1) Does it detect prostate cancer? The answer is "sometimes," but further refinement of this ability is desirable.

2) Does it detect prostate cancer reliably enough? The answer to this depends on what decision hangs in the balance: reliable enough to justify referral for TRUS? For radical prostatectomy? For making underwriting decisions on a life insurance application?

3) Does it detect prostate cancer early enough? The ultimate correlation between PSA information in some form (PSA level, PSAD, PSAR, PSA+something else) and localized curable prostatic cancer has yet to be determined.

Table 1
The Three Abstracts in this Issue

<table>
<thead>
<tr>
<th>Abstract</th>
<th>Subject</th>
<th>Entry Period</th>
<th>lo</th>
<th>E</th>
<th>Mean</th>
<th>Range</th>
<th>FU Completeness</th>
</tr>
</thead>
<tbody>
<tr>
<td>350L1</td>
<td>Angina, Mayo Clinic</td>
<td>1927-1944</td>
<td>599</td>
<td>537,480</td>
<td>6.3</td>
<td>5-23</td>
<td>*</td>
</tr>
<tr>
<td>350L2</td>
<td>Angina, Paul Dudley White</td>
<td>1920-1931</td>
<td>456</td>
<td>4787</td>
<td>10.5</td>
<td>14-36</td>
<td>99%</td>
</tr>
<tr>
<td>158M1</td>
<td>Pancreatic Ca, multinational</td>
<td>1946-1989</td>
<td>2015</td>
<td>15,003</td>
<td>7.4</td>
<td>&lt;2-25+</td>
<td>*</td>
</tr>
</tbody>
</table>

* few were lost to FU, but a large percentage were withdrawn alive at end of study.
Table 2

Risk Factors for Prostate Cancer

1. Age

<table>
<thead>
<tr>
<th>Age</th>
<th>Incidence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-59</td>
<td>15-25%</td>
</tr>
<tr>
<td>60-69</td>
<td>20-30%</td>
</tr>
<tr>
<td>70-79</td>
<td>30-45%</td>
</tr>
<tr>
<td>80+</td>
<td>50-70%</td>
</tr>
</tbody>
</table>

(* incidence of "latent" or indolent prostate cancer, based on autopsy studies. Scardino, PT. Early detection of prostate cancer. Urol Clinics of No Amer 1989;16:642)

2. Race

Age adjusted incidence* of prostate cancer (U.S.) by race and year of diagnosis.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>black</td>
<td>105.1</td>
<td>125.5</td>
<td>136.1</td>
</tr>
<tr>
<td>white</td>
<td>62.4</td>
<td>78.0</td>
<td>99.2</td>
</tr>
</tbody>
</table>


3. Positive Family History

<table>
<thead>
<tr>
<th>Affected Relatives</th>
<th>Increased Prostate Cancer Risk (PCA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>one relative w/PC before age 65*</td>
<td>4x</td>
</tr>
<tr>
<td>one first degree and one second degree** relative</td>
<td>8x</td>
</tr>
</tbody>
</table>

(* Scardino PT. ibid. p 651.)


A first degree relative is a brother or father and second degree relative is an uncle or grandfather.

Table 3

Mortality rates per 100,000 for prostate cancer (1986)

<table>
<thead>
<tr>
<th>Age group</th>
<th>U.S.</th>
<th>W. Europe</th>
<th>E. Europe</th>
<th>Nordic</th>
<th>East Asia</th>
<th>Oceania</th>
</tr>
</thead>
<tbody>
<tr>
<td>45-54</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>55-64</td>
<td>25</td>
<td>19</td>
<td>21</td>
<td>25</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>65-74</td>
<td>111</td>
<td>102</td>
<td>107</td>
<td>132</td>
<td>23</td>
<td>116</td>
</tr>
<tr>
<td>75-84</td>
<td>322</td>
<td>326</td>
<td>289</td>
<td>421</td>
<td>72</td>
<td>355</td>
</tr>
<tr>
<td>Total</td>
<td>68</td>
<td>65</td>
<td>63</td>
<td>84</td>
<td>15</td>
<td>72</td>
</tr>
</tbody>
</table>

Percent Annual Change in Mortality Rates from 1969 through 1986, all ages combined (%/yr)

<table>
<thead>
<tr>
<th>U.S.</th>
<th>W. Europe</th>
<th>E. Europe</th>
<th>Nordic</th>
<th>East Asia</th>
<th>Oceania</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.9</td>
<td>0.7</td>
<td>1.3</td>
<td>2.3</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Table 4
Cost/Benefit Formula and Variables, for Screening Tests*

\[ V = p n w B + (1-p) c R - S - [p n + (1-p)(1-c)] w (T + A) - (1-p)(1-c) F \]

<table>
<thead>
<tr>
<th>Value Assigned</th>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$29,443</td>
<td>V</td>
<td>net value or benefit per individual screened</td>
</tr>
<tr>
<td>$90</td>
<td>B</td>
<td>benefit from early treatment</td>
</tr>
<tr>
<td>$90</td>
<td>R</td>
<td>reassurance value of a negative result</td>
</tr>
<tr>
<td>$20,000</td>
<td>S</td>
<td>screening cost per test</td>
</tr>
<tr>
<td>$1413 to $1477</td>
<td>T</td>
<td>treatment cost</td>
</tr>
<tr>
<td>$1413 to $1477</td>
<td>A</td>
<td>adverse reaction cost</td>
</tr>
<tr>
<td>$1413 to $1477</td>
<td>F</td>
<td>false positive cost</td>
</tr>
</tbody>
</table>

Probability

<table>
<thead>
<tr>
<th>Probability</th>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>.053</td>
<td>p</td>
<td>prevalence of target disease in screened population</td>
</tr>
<tr>
<td>.67-.79</td>
<td>n</td>
<td>sensitivity of screening test</td>
</tr>
<tr>
<td>.93-.59</td>
<td>c</td>
<td>specificity of screening test</td>
</tr>
<tr>
<td>.85</td>
<td>w</td>
<td>willingness: probability of returning for treatment</td>
</tr>
</tbody>
</table>

(* formula developed by Godderis and Bronken\textsuperscript{19} for general screening model, and modified by ACS Prostate Project\textsuperscript{20})

Table 5
Possible Uses of PSA as a Test for Prostate Cancer

SCREENING

1) PSA in stand-alone screening
   Stand-alone measuring referenced to action-point
   a) one-time PSA level (or isomers?) population norms ?significant elevation
      PSA density (PSAD) prostate gland size ?abnormal ratio
   b) serially PSA rate of change (PSAR) self-standardized ?significant interval rise or disturbing trend

2) PSA in conjoint screening
   Conjoint involving action-point
   a) before or after DRE certain combinatorial rules ?either/both abnormal
   b) with TRUS certain combinatorial rules ?either/both abnormal
   c) DRE TRUS certain combinatorial rules ?either/both abnormal
   d) other

3) PSA as a confirmatory test
   If sensitivity and specificity were ever good enough, and pre-test probability ever high enough, could PSA be "reliable" enough to diagnose pancreatic cancer?

DIAGNOSTIC

4) PSA for recurrence detection
5) PSA for progression detection, in "watchful waiting" situations.
Table 6

Combinatorics of Sensitivity (SN) & Specificity (SP)

<table>
<thead>
<tr>
<th>Test</th>
<th>SN</th>
<th>SP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A: PSA (10)</td>
<td>.80</td>
<td>.60</td>
</tr>
<tr>
<td>Test B: DRE</td>
<td>.50</td>
<td>.80</td>
</tr>
<tr>
<td>Both A and B</td>
<td>.40</td>
<td>.92</td>
</tr>
<tr>
<td>Either A or B</td>
<td>.90</td>
<td>.48</td>
</tr>
</tbody>
</table>

If decision requires both tests to be positive,
combined SN = \( SNA \times SNB \)
combined SP = \( SPA + (1-SPA)SPB \)

If decision requires EITHER test to be positive,
combined SN = \( SNA + (1-SNA)SNB \)
combined SP = \( SPA \times SPB \)

The predictive value of the combined tests results can then be derived by using Bayes' formula with the combined SN and combined SP.

References