REDUCING DIAGNOSTIC ERROR IN MEDICINE – IN RESEARCH

PAINTING BY THE NUMBERS: RESEARCH RELIANCE ON A SINGLE MEASUREMENT LEADS TO ⬆ ERROR

© 2016 Mark Gusack, M.D.
MANX Enterprises, Ltd.®

1. Research purporting to show correlation between single quantitative test results and a diagnostic decision point is often flawed. 2. The tendency to place too much significance upon numbers needs to be counterbalanced by refocusing attention on the patient's unique clinical circumstances. 3. A major sea change needs to take place in the culture of medicine to counter the adverse effects of an increasingly innumerate profession buried in a blizzard of statistical analyses.

APPLES TO ORANGES TO BANANAS

At many hospitals the pathologist is routinely asked by the orthopedic surgeon to count neutrophils on intraoperative frozen sections of tissues taken from a patient's knee to determine if their failing prosthetic device is infected.

However, there is a small problem:

Mirra JM et. al. concluded in 1976 that ≥ 5 neutrophils/HPF has **high sensitivity** for predicting infection in loosened prosthetic devices at high power magnification.[1]

Yet Lonner, JH, et. al. concluded in 1996 that ≥ 5 neutrophils/HPF has **low sensitivity** for predicting infection in loosened prosthetic devices at high power magnification.[2]

QUESTION

WHY IS THERE A DIFFERENCE BETWEEN THE TWO STUDIES?

ANSWER I

Klein notes that:[3]

"Mirra et al used a high-power objective of 50x whereas most other authors cited in the remaining studies used a 40x objective".

He goes on to state:

"Unfortunately, while each study may be internally valid, there is to date none that also reports the field diameter of the ocular used on the same microscope. The only thing that 40x objectives on different microscopes have in common is that if the same power ocular is used on each, the cells will appear the same size in each microscope."

Real world pathologists use 40x invalidating the first study completely!

ANSWER II

A superficial evaluation of the statistical methods of Mirra et al. reveals significant problems:

- ➡ They do not use PMN count directly in their statistical analysis.
- ➡ Instead, they rank the counts 0 to 3+ into very unequal groups as shown next:
- ➡ 0 = 0 PMN's, 1+ = 1-5 PMN's, 2+ = 6-49 PMN's, and 3+ = 50 and more PMN's
- ➡ They do not provide any justification for this unusual set of groupings.
- ➡ And even worse…Mirra et al do not base their work on frozen section material!

And that's just the beginning. Take a look at a few more potential bias's that are not addressed listed on the other side of this poster. Yet the journal editor allowed this to be published? And no one questioned it's conclusions, let alone it's methodology?

EXAMPLE OF THE EFFECT OF FALSE ASSUMPTIONS

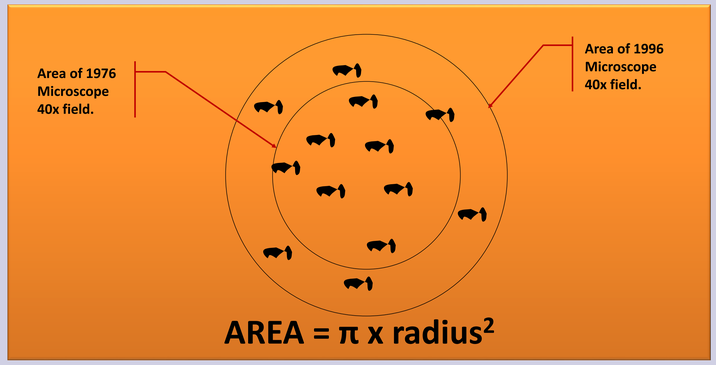
The most critical false assumptions made by the authors **and their readers**:

- ➡ 50x was the high power to be used despite the fact that pathologists use the 40x objective!
- ➡ That magnification specifies the cross sectional area and image quality
- ➡ That these features have not changed during the 20 year period between the two studies

The truth:

- ➡ The cross sectional area is defined by the quality of the optics and varies between microscopes
- ➡ The cross sectional area at 40x rose steadily from 1976 to 1996
- ➡ The image quality image rose as well during this time making it easier to identify neutrophils

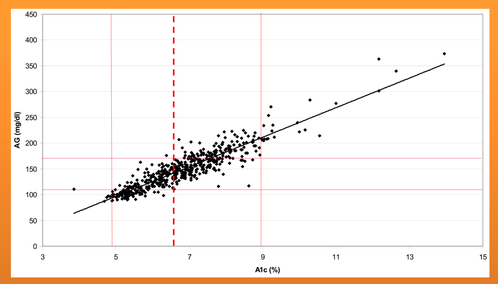
GRAPHIC EXAMPLE OF FALSE ASSUMPTIONS MADE



It turns out an approximate 50% increase in the average **radius of view** of the 40x objective over 20 years has caused the **area of view to increase by a factor of 225%**. Now even 10 neutrophils per high power field is not adequate to differentiate between infection and no infection.

ANOTHER MINI-EXAMPLE OF MINDLESS APPLICATION OF NUMBERS

WE ARE CALCULATING ESTIMATED AVERAGE GLUCOSES FROM THIS?



Linear regression of A1C at the end of month 3 and calculated AG during the preceding 3 months.
Calculated $AG_{mg/dl} = 28.7 \times A1C + 46.7$ *(AGmmol 1.59 A1C 2.59)* (R^2 **0.84**, P *0.0001)*[4]
The R value of 0.84 is completely unacceptable for clinical application in the single patient

IT'S EVEN WORSE FOR ESTIMATED GLOMERULAR FILTRATION RATES CALCULATED FROM SINGLE CREATININE RESULTS!

REFERENCES – SELECTED:
1. Mirra JM et. al. **The Pathology of the Joint Tissues and its Clinical Relevance in Prosthesis Failure**; Clinical Orthopedics; 117:221-240 1976.
2. Lonner, JH, et. al. **The Reliability of Analysis of Intraoperative Frozen Sections for Identifying Active Infection during Revision Hip or Knee Arthroplasty**; The Journal of Bone and Joint Surgery 78-A:10 1553-1558 Oct 1996.
3. Klein, M. **Letter to the Editor** American Journal of Clinical Pathology 2009; 131:147
4. DM Nathan et al; **Translating the A1C Assay Into Estimated Average Glucose Values;** Diabetes Care. 31: 1473-8 2008.
ACKNOWLEDGEMENTS: Lionel Stephens 10th grade mechanical drawing teacher Bethesda Chevy Chase HS and William S. Yamamoto, M.D. director Clinical Engineering GWU Medical School

OUTCOME OF MY INVESTIGATION

I developed a presentation illustrating the reasons why making a neutrophil count intraoperatively at frozen section on randomly submitted tissue was inappropriate. This presentation was shown repeatedly to our orthopedic surgeons for six years.

The results? Neutrophil counts were ordered for every joint replacement operation despite the fact that the results of my investigation are definitive and unassailable.

This means that:

- ➡ Patient safety was compromised if infection was present but not identified
- ➡ Patient quality of life was affected by both false positives and false negatives
- ➡ Valuable resources were squandered at every surgical procedure when ordered

A CASE OF SYSTEMATIC BIAS

BELOW IS A PARTIAL LIST OF QUESTIONS NOT ANSWERED BY EITHER RESEARCH PAPER:

- ➡ WERE ALL REVISIONS OF PROSTHESES SAMPLED FOR NEUTROPHILS?
- ➡ IF NOT, WHAT CRITERIA WERE USED BY EACH SURGEON?
- ➡ WHAT WERE THE VARIOUS TYPES OF PROSTHETIC DEVICES?
- ➡ WERE TISSUES SUBMITTED SIMILAR IN NATURE AND FROM THE SAME SITES IN THE OPERATIVE FIELD?
- ➡ WHAT, IF ANY, ANTIBIOTICS WERE USED PRIOR TO SURGERY? AFTER SURGERY?
- ➡ WHAT DECISIONS WERE MADE BASED ON EACH FROZEN SECTION RESULT?
- ➡ WHAT WERE THE SHORT AND LONG TERM OUTCOMES IN TERMS OF FALSE POSITIVES/FALSE NEGATIVES?
- ➡ WHAT WERE THE ORGANISMS CULTURED AND FROM WHAT TISSUES?

In the literature it appears that the rate of infection is going up and the types of organisms causing infection are changing. This leads to other questions like:

- ➡ HAVE NEWER PROSTHESES CHANGED INFECTION/INFLAMMATION/FAILURE RATES?
- ➡ HAVE CHANGES IN SURGICAL TECHNIQUE CHANGED INFECTION/INFLAMMATION/FAILURE RATES?
- ➡ IS PMN COUNT DIFFERENT FOR DIFFERENT BACTERIAL SPECIES CAUSING INFECTION?

THEN THERE IS THE PATHOLOGY SIDE OF THIS "EQUATION."

- ➡ WHAT WERE THE METHODS OF GROSS EXAMINATION INTRAOPERATIVELY?
- ➡ WHAT WAS SAMPLED AND WHAT VOLUMES OF TISSUE WERE FROZEN?
- ➡ DID THE PATHOLOGIST DOING THE EXAM HAVE ADEQUATE EXPERIENCE?
- ➡ WHAT WAS THE METHOD OF FREEZING AND CUTTING SPECIMENS?
- ➡ WHAT WAS THE NUMBER OF SLIDES CUT FOR EACH TISSUE BLOCK?
- ➡ WHAT WERE THE METHOD OF STAINING AND COVER SLIPPING SLIDE?
- ➡ WAS BIASING CLINICAL INFORMATION AVAILABLE TO THE PATHOLOGIST
- ➡ WHAT WAS THE INTERNAL CONCORDANCE FOR EACH READING PATHOLOGIST?
- ➡ WHAT WAS THE CONCORDANCE BETWEEN READING PATHOLOGISTS?
- ➡ WHAT WAS THE ACTUAL METHOD OF COUNTING NEUTROPHILS?

A review of the literature shows a wide variation in sensitivity, specificity, positive and negative predictive power as well as wide variation in percent cases positive for PMN's making comparison of data across articles impossible. I found that the pathology literature mostly discounts the usefulness of counting PMN's despite "firm" support in the clinical literature.

CONCLUSION

- ➡ The use of poorly defined measurements and the inappropriate application of statistical methods is a problem that adversely affects the healthcare field.
- ➡ The clinical and laboratory phenomena we observe cannot be described entirely by a single "made up" quantitative result taken out of context.
- ➡ Instead, these phenomena must be carefully defined and fully characterized before we can use them, through the application of **The Scientific Method**, to:

REDUCE DIAGNOSTIC ERROR IN MEDICINE