Workload in Gynecologic Cytology: Setting the Bar Higher

Dr. Andrew Renshaw
Baptist Hospital
Miami, FL
Special thanks to:

- Manon Auger, MD
- Tarik M. Elsheikh, MD
- Heather Wilgenbusch, CT (ASCP)
- Nick Dudding
Outline

- Why study workload?
- Tools for studying workload
- Workload data
- Summary
- Where to we go from here?
Why study workload?

- It’s the law

- CLIA 88 requires every cytology director to evaluate the workload of every individual CT at regular intervals

- Obviously after 20 years we must be pretty good at it - what’s the problem?
No one trusts other labs
TRUST

- What QA data do you need to trust a lab enough to only get a diagnosis and the QA data on your loved ones Pap?

- For me, a test must have documented:
  - Accuracy (Sensitive and Specific)
  - Reproducibility (precision)
  - Clinically relevant/usefulness
What do we document?

- Specificity
  - Biopsy fu, HPV + rate

- Clinical utility

- Do not document
  - Sensitivity
  - Precision

- So what?
Why is this important?

- High workload labs
  - Lowest cost (140 slides/day)
  - Best QA data in the world: error rates of <1 in a million
Conclusion

- Either all labs should copy the high workload labs
  
  or

- There is something wrong with the QA
What’s wrong with the QA?

- Reliance on insensitive surrogates
  - Naryshkin - 10% review of negative slides
  - Boone - infection pick up rate
  - LSIL

- Lack of controls - can’t measure sensitivity without it – would never run a chemistry analyzer without them
Can you treat CTs like a chemistry analyzer?
–NO–
much more variable

- need better and more creative QA measures to account for CTs not being machines, but you HAVE to measure their performance

- Controls are not optional for good laboratory practice in an industrial setting
Tools for studying workload

1. Measure CTs not the Lab
2. Controls – pre-screen
3. Surrogates
4. Forget LSIL
5. Precision
6. Too much atypia?
1. Measure CTs

- The sensitivity of the Pap smear rests ONLY in the hands of the CTs
- I don’t care how good the lab is, I care how good the CT who reads my slide is
- CTs vary considerably
- Most labs can not measure FNP, ASC/SIL ratio etc for CTs, only lab
2. Controls

- CAP Checklists
- CP: >50 questions about controls
- AP: 13 questions about controls
- Cytology: 0 questions about controls

- Fundamental laboratory medicine
- Not possible to measure sensitivity without controls
False Negative Proportion

- Uncontrolled test
- Can get excellent results by doing a lousy job of re-screening
- Always under-estimates error:
  - FNP 0-2% vs 20-40% by controlled studies
Pragmatic test

- Pragmatic test: no useful studies using 10% re-screening and workload over last 20 years
Controls in Gyn Cytology

- Seeding – logistics, controls always too easy: Australian experience

- Blinded pre-screening – routine speed is fine

- Time division - Tarik’s method
Advantages of pre-screening

- Controlled
- Can measure sensitivity
- The reported overall sensitivity matches results in FDA trials (not all 99%)
- Detects as many errors as Focalpoint
- Improves sensitivity of routine screening (better feedback)
- Improves reproducibility of routine screening (precision) (better feedback)
- Improves sensitivity or pre-screening (better feedback/more time)
Pre-screening is the only choice for real life QA

Remaining questions?

- How fast? – 30 vs 120 seconds vs full (best yield and possible bias)
- Does it work for automated screening?
- How many slides?
- How best to implement it
Why people won’t pre-screen

- CLIA says you don’t have to
- CLIA 88 says re-screen
- Workload limits
- Math is too complicated

- Is there anything else than pre-screening?
Dividing the day

- AM vs PM
- Days of the week
- 4 time periods

- Can’t measure sensitivity, but can show that one time is better than another
3. Surrogates

- Abnormal rate, ASC rate, ASC/SIL ratio, infection rate, FNP all correlate with sensitivity but:

- Correlation stronger in poor labs with variety of screeners

- Worthless in labs where most CTs are the same (either good or bad)
Pragmatic test

- Only works if you have one or two CTs with sensitivity of 30-40% less than the rest

- Not a very high bar - is a sensitivity of 70% really OK?
4. Forget LSIL

- Really interested in sensitivity for HSIL – but too rare to measure (under-powered)

- ASC or LSIL?

- “LSIL is a more reproducible (RIGHT) and therefore more clinically relevant, and more accurate surrogate for the sensitivity of HSIL than ASC” (WRONG)
ASC vs LSIL

- Clinical relevance
- Practical utility
- Administrator’s dream
- Pragmatist’s view
Clinical relevance

- Clinical: Risk of CIN 2-3 the same for ASC and LSIL (10%)
- One third of all CIN2-3 first detected by an ASC Pap
- ASC is what makes the Pap test work
Practical utility

- Where it has been examined, the performance at HSIL is more closely correlated with the performance at ASC than LSIL.
- People miss ASC and ASC-H and HSIL, they don’t miss LSIL - most LSIL is easy.
- If you use LSIL every mistake becomes ASC.
- *TIS at high speeds good at LSIL lousy at ASC and HSIL.*
Administrator’s dream

- There is a reason the commercial labs use LSIL as their threshold
  - Errors are rare
  - Errors you do find are easy to call ASC
  - Can go at high speeds and not miss LSIL (while missing ASC and HSIL)
Pragmatist’s view

- Studies using LSIL exist, and every lab is excellent

- Not very useful, but...
The pragmatist’s result:

- If you insist on using LSIL as your threshold, you might as well buy the TIS system, do 400 slides a day, and have a manual review rate of 0%

- Your QA results will look fine
5. Precision

- Gyn cytology is not very precise
  - Abnormal rate varies 30% within labs
  - ASC/SIL ratio varies 600% within labs
  - Can’t both be as good
Precision

Feedback from supervisor, pathologists, other CTs

- Improvement
  - LGS – biggest effect on ASCUS
  - Rapid pre-screening
6. Too much atypia?

- ASC/SIL ratios vary widely
- LGS (Thinprep) often increases ASC
- ASC is “uncertain”
So?

- Reflex HPV testing can account for specificity (95%) so cytology should focus on sensitivity.

- Can construct ROC curves with sensitivity versus women sent for colposcopy.

- Results are very different with and without HPV testing.
Optimum ASC/SIL Ratio

- No HPV testing: 1 or lower
- With HPV testing: 2 – 2.5
Workload Data

- What factors affect my performance at different workloads?
- How fast can a CT go manually?
- How fast can a CT go using TIS?
Time and workload

- Tarik’s data
  - Days of the week
  - AM vs PM
  - How many hours a day
Days of the week

- 4 CTs screening 50-85 slides/day with TIS

- 3/4 detected significantly less abnormal cases on Monday or Friday

- Should CTs screen less on these days (CLIA 88 individualized limits)?
AM vs PM

- 4 CTs screening 50-85 slides/day with TIS
- 2 of 4 CTs detected significantly less abnormal cases in the PM vs the AM
- 1 CT detected the same # of abnormal cases, but in the afternoon they were all classified as ASC
- Should CTs screen in the afternoon? (CLIA 88 individualized limits)
How many hours a day?

- 10 CTs doing 40 - 90 slides/day with TIS

- 9/10 screened significantly more slides, were more sensitive, and more specific during the first two hours

- The 10th CT was a supervisor who did lots of other work

- How many hours should CTs screen (UK 4 hours)?
These results are a challenge for cytology directors, but

Most CTs who are poor screeners can do well for a limited number of hours

If we get workload right we might be able to improve the performance in the lab (especially since cytology directors seem adverse to firing CTs)
How fast? manual
Manual: conventional smears

Figure 2. Individual (n = 22) and Laboratory (n = 8) Sensitivity and Workload, Manual Screening

![Graph showing sensitivity and workload relationship](image)
Manual screening of SurePath

\[ y = -0.1905x + 101.42 \]

\[ R^2 = 0.1458 \]

sensitivity vs. slides/day graph.
Manual

- Conventional and Surepath similar

- >30 slides/day, no chance at 100% sensitivity
How fast: Automated?

- Thinprep Imaging System (TIS)
- No useful data on FocalPoint (3 pts, all same workload)
- TIS vs manual two ranges
- Is 100 the right threshold?
- ECA adjusted workload
TIS vs manual

- 2 ranges
  - 60-100 slides/day
  - >100 slides/day
60-100 slides/day with TIS vs manual

- 9 studies
- HSIL increased in all 9: 13-42% (improved sensitivity)
- When biopsied: all showed same or greater rate of CIN2-3
- >100 slides/day TIS is better than manual for HSIL
> 100 slides/day

- 5 studies
- No increase in HSIL
- At 140 slides/day 330% increase in FNP
- 1 exception 80 - 140 slides increased HSIL
- TIS has a problem with HSIL >100 slides/day
Is 100 the right threshold?

- Tarik’s lab

- 8 years of individual data for up to 11 CTs – big numbers

- Measured ECA, ASC and HSIL
Going from 67-76 slides/day: no change in performance

Going from 76 to 80 slides/day: HSIL rate decreased (.9% to .6%)
Going from 80 – 90 slides/day:
  • Decreased ASC and HSIL in some CTs

Going from 90 -118 slides/day:
  • Decrease in ASC and HSIL in all but one CT
Conclusion

- In Tarik’s lab, some CTs start to fail at 70 slides/day, most fail at 100 slides/day
Automated screening: can’t you just graph it?

- Yes but:
- Workload alone not useful
- ECA adjusted workload: workload x %
  ECA
Figure 4 Laboratory Sensitivity and ECA Adjusted Workload for ThinPrep Trial and Raab
How fast for my lab?

- 700/ECA rate (10 for 10%)
- Tarik’s ECA rate 10% - 70 slides
- A lab with an ECA rate of 5% can screen 140 slides a day (though sensitivity for HSIL above 100 slides/day not known)
- If ECA rate rises, TIS loses value (HPV screening)
Summary

- Evidence based evaluation of workload in the cytology laboratory generates a very different group of QA measures than most laboratories are currently measuring.
Recommendations non workload

- Measure clinically relevant values:
  - CTs not lab
  - Errors with ASC not LSIL (or do 400 slides/day)

- Measure and Improve sensitivity
  - pre-screening

- Improve precision with
  - ASC/SIL ratios between 2-2.5
  - LGS
  - pre-screening
Recommendations: workload general

- Manual CTs should do no more than 30 slides/day

- TIS CTs should do no more than 700/ECA rate slides/day or 100 slides/day (for HSIL)

- Should consider different limits for Monday and Friday, AM vs PM

- Should consider screening < 8 hours/day
Recommendations workload: exceptions

- Not all CTs should do the same workload (CLIA 88 individual limits)

- Allow individual CTs to go faster if they document performance
  - Pre-screen
Future directions: what questions should we be asking?

- Patients
- Cytology medical directors
- Cytotechnologists
Patients

- Do you trust your lab?
  - Accuracy (Sensitivity and Specificity)
  - Reproducible (Precision)
  - Clinically relevant (CT not lab performance)

- Why not insist on a chance at 100% sensitivity?
- Would you be willing to pay more for 100% sensitivity?
Cytology medical directors

- What sensitivity am I aiming for?
- How am I measuring it?
- What controls am I using to ensure I am measuring it correctly?
- Am I treating each CT individually?
Cytotechnologists

- How can I avoid being forced to screen faster than I feel comfortable with?

- How can I avoid being locked in a sound proof room at a commercial lab?
The answers are all the same!

- Use controls
  - Pre-screen
  - Set workload based on controlled data

- Doing QA well in this environment is an art as well as a science
Will this happen? Probably not

- Using controls, measuring performance with methods that work, and setting limits for acceptable performance were all developed over 50 years ago: standard good laboratory practice in medicine and industry

- Cytology directors won’t touch controls:
  - What they say: I am too busy to do more than CLIA demands
  - What I hear: I am too busy signing out cases to worry if I am missing cancer or not
The irresistible allure of productivity

- 4 FDA trials over 20 years
  - Control arm sensitivity 80%
  - Control arm used to be conventional, now its LBC – what happened?

- Every lab is working faster

- We are losing the gains of technology in the search for productivity and the QA measures directors rely on are too insensitive to catch it
The future of gynecologic cytology

- Where does business go?
  - Politics
  - Cost
  - Quality

- In the absence of more useful QA measures, all the business will end up at the commercial labs

- Our only defense is to set the QA bar for workload higher

- Making a market for higher quality is up to us!