Implications for AML Raised by “Chemo-free” ATO+ATRA in APL

- Doing trials in patients with high likely success rates
- Doing trials in small subsets
- MRD

Ethical Dimension of ATO+ATRA w/ o “Chemotherapy”

- AIDA cures 90-95%
- Without new trials cure rate will never be “100%”
- But do the ends justify the means?
- Cost of unsuccessful trial can be quantified
- Decisions to proceed can be based on cost

Quantifying Cost

- Cohorts of 6: If >90% prob. CR rate < 90% stop accrual
- If true CR rate 60% trial stops w/median 12 pts.
- 12 pts. X 60% CR = **7 CR** vs. with AIDA 12 pts. X 90% = **11 CR**
- **Potential cost = 11-7 = 4 patients**
- If true CR rate < 60% trial stops earlier: same cost
- Oversimplification if accrual quicker than patients can be evaluated
- High false negative rate must be accepted

Adaptive Randomization

- Bayesian rather than group sequential design
- Use interim data to **repeatedly** compute probability that one Rx is "better" than the other(s)
- Unbalance randomization to favor better Rx
- Closed arm can re-open

Nature Reviews Drug Discovery 2006;5:27-36
JCO 2003;21:1722-27

Disadvantages and advantages of ARAN

- Bayesian prior probs. subjective (but so are p-values)
- Small sample sizes mean higher type 1, type 2 errors
- Unequally sized groups do likewise
- Feasibility: time to observe outcome vs. accrual rate, computing
- Loss of biologic information
- **Operates in accordance with how patients think doctors practice (or should practice)**

Statistical Analysis and the Illusion of Objectivity. American Scientist 1988;79: 159-65

The P-value Fallacy Ann Intern Med 1999;130:995-1004

“Molecularization” of AML

- More subgroups
- With same treatment response may differ by subgroup
- Subgroups may be treated differently
- 2-sided p=0.05, power =80% to detect small difference not become feasible: not enough pts of a particular type

Accounting for Heterogeneity In Response to Same Rx

- One extreme : ignore it (as in Simon 2-stage = S2S)
- Second extreme : separate trials for each subgroup – can't use data from 1 trial to adaptively affect conduct of others
- Third method: consider subgroup-treatment interactions (STI); adaptively use data to see to what extent subgroups can be combined ("borrowing strength")

Stat Med 2008; 27:2802-15

Simulations of STI vs. S2S for Drug X

Subgroup	True CR Rate	Probability (Reject X) S-TI	S2S	Mean # Pts S-TI	S2S
Better	***0.58***	0.10	0.75	21	10
Worse	0.11	0.90	0.75	19	25

S2S historical 0.2 , goal 0.4, type 1 error = .10, type 2 =. 20;
¾ pts in worse group

Problems with Conventional Phase II → Phase III Setup

- Wastes information : single-arm phase II survival data

  can't be used in phase III

- Even w/randomization in phase 2 decision to go to phase 3 rests on response data, assuming correlation between response & survival

- Delay between phase 2 and phase 3

Seamless P 2/P 3

- Randomizes to E or S throughout
- Repeated interim decisions based on response *and survival and relation* between these (mixture model)
- Possible decisions:
  - stop, conclude E better
  - stop, conclude E no better (begin new E)
  - continue trial
  - expand trial (phase 3 begins without interruption accrual)
- Simulations show shorter trials with fewer pts. compared to Simon 2-stage followed by phase 3 group sequential design & no increase in type 1 or 2 errors (mixture model)

Biometrics 2002;58:823-31

MRD

- Morphology lacks sensitivity → routine marrows not recommended in f/u

  (Estey & Pierce Blood 1996:87:3899-3902)
- More sensitive methods to detect MRD now available
  - MFC 0.1% - 0.01% (vs 5% for morphology): applicable to all pts.
  - FISH 0.2% : applicable to 2/3s pts.
  - PCR/NGS 0.01-0.001 %: applicable to 2/3s pts.
- High positive predictive value for morphologic relapse
- Short interval between MRD & morphologic relapse
- More predictive value than pre-Rx covariates but predictive still limited with single MRD determination
- Value in assigning therapy?
- Does reducing MRD delay morphologic relapse?
- Standardization?
- Is there remaining need for morphology?

Associations with relapse in t(8;21)

	HR (95% CI)
Log WBC pre Rx	2.10(0.73-6.04)
RTK mutation pre Rx (CKIT or FLT3)	1.51 (0.65-3.54)
MRD reduction ≥ 3 logs	0.24 (0.10-0.57)

Jourdan et al. Blood 2013;121: 2213-2223 (all received HDAC, none received GO, all age < 60)

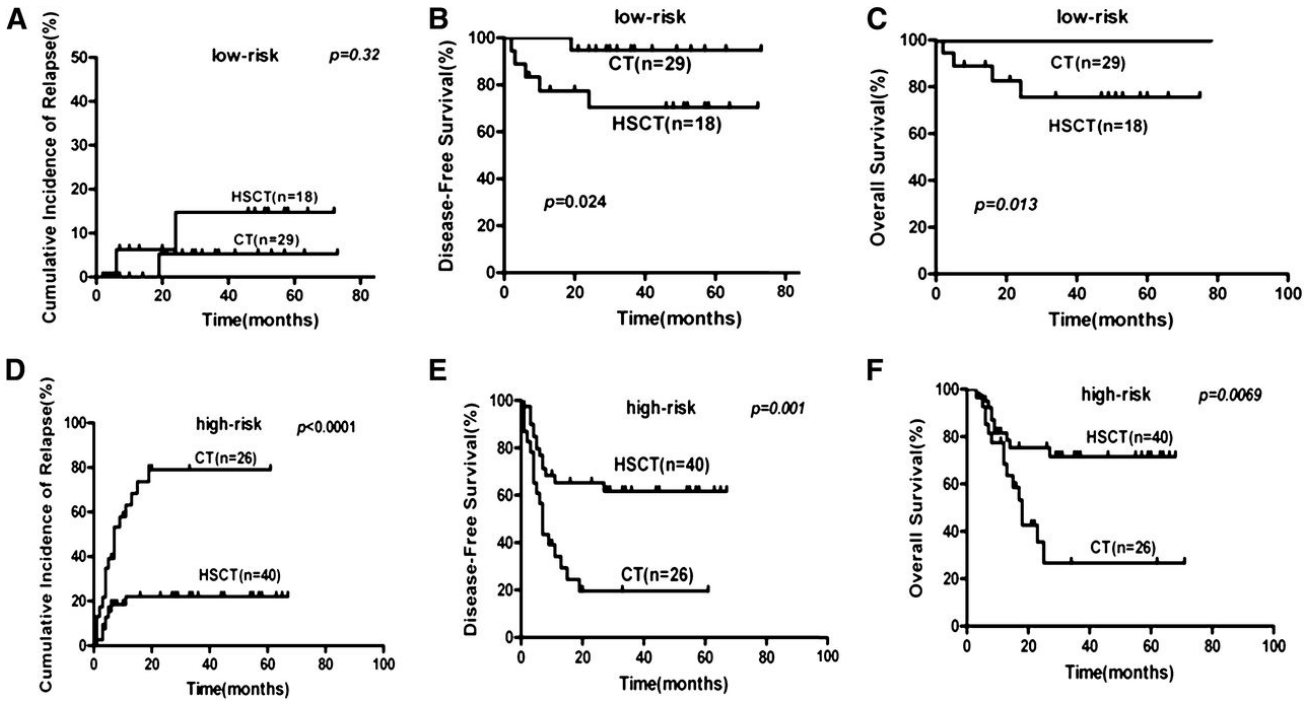
How Much Does Addition of MFC Data Improve Prognostication?: SWOG SO106

	C-statistic	
Covariates	RFS	Survival
Basic	0.61	0.63
Basic+cyto.	0.63	0.66
Basic+cyto +NPM/FLT3 ITD	0.65	0.69
Basic+cyto+NPM/FLT3 + MFC	0.66	0.70

C-statistic 0.6-0.7 poor, 0.7-0.8 fair, 0.8-0.9 good ability of a model to forecast

Walter et al. Leukemia 2016; 30: 2080-83

t (8;21) :MRD +(high risk) plan HCT, MRD neg (low risk) no HCT



Hong-Hu Zhu et al. Blood 2013;121:4056-4062

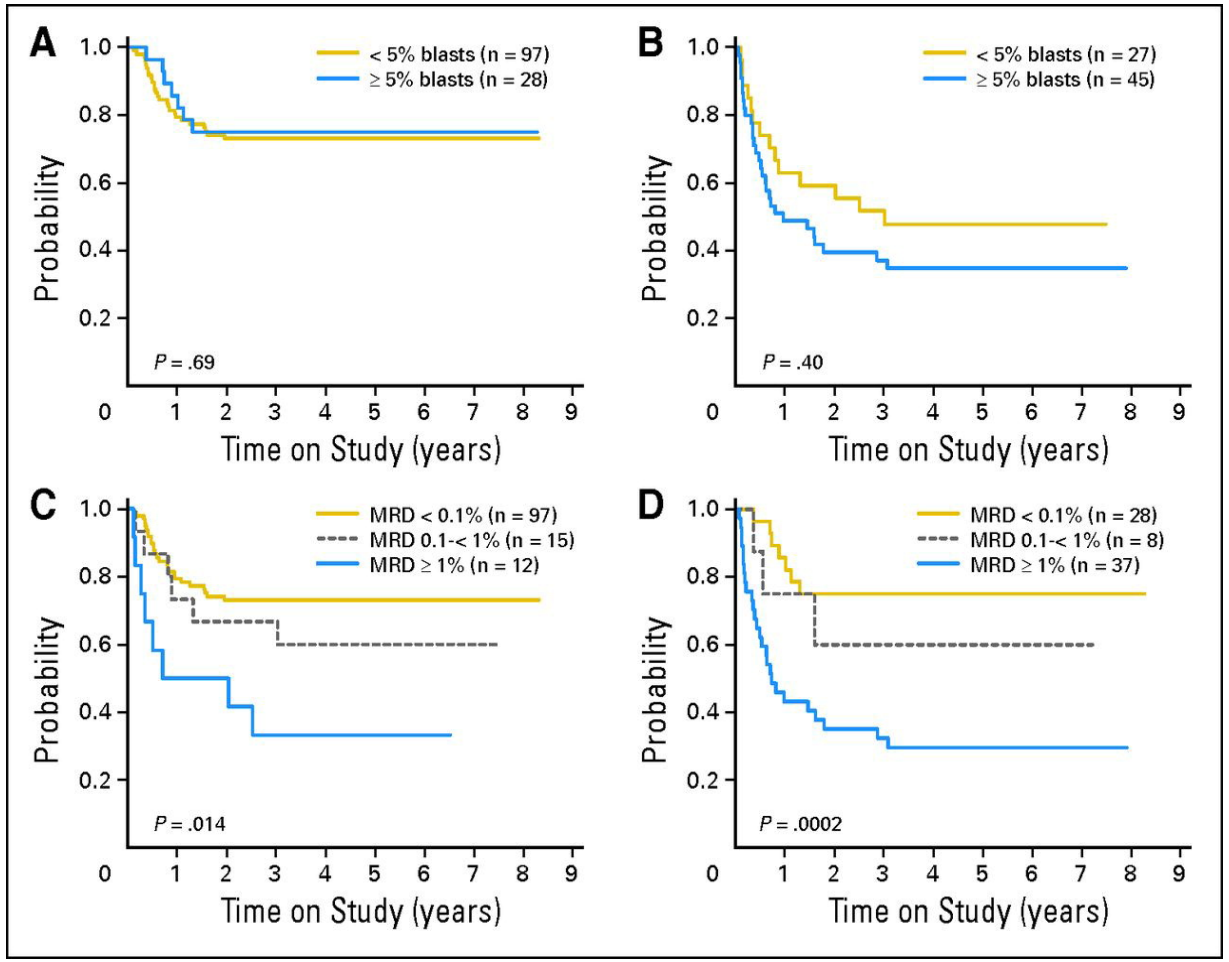
Value of Reduction of MRD

- Randomize between new Rx when MRD detected vs. only when morphologic relapse detected, e.g. should HCT be delayed to Rx MRD?
- Requires availability more new Rxs. to treat MRD; rare per clinicaltrials.gov
- ELN recognizes CR w/o MRD as distinct→ implies relapse with MRD only is also distinct
- Will reduce time to relapse vs. current morphology-based criterion(bad for drug approvals)
- May not bring clinical benefit: rationale for NCCN recommendation against routine marrow f/u
- But might facilitate discovery of new drug activity & be associated with favorable benefit/risk depending on Rx chosen : HCT vs. less intense
- AML 18:randomize between MRD monitoring & not in CR

Need morphology to detect relapse?

- 96 cases of relapse (>5% blasts by morph.)
- 95 were MFC(10-color) pos.; 1 suspicious
- False negative rate 0/96 (95% CI 0-4%)
- 357 cases with <5% blasts by morph
  - 254 were MFC neg., 103 MFC pos.
- ***Nominal*** false positive rate js 103/357 (29%)

Zhou et al. Leukemia in press


A
< 5% blasts (n = 97)
≥ 5% blasts (n = 28)
Probability
P = .69
Time on Study (years)
B
< 5% blasts (n = 27)
≥ 5% blasts (n = 45)
Probability
P = .40
Time on Study (years)
C
MRD < 0.1% (n = 97)
MRD 0.1-< 1% (n = 15)
MRD ≥ 1% (n = 12)
Probability
P = .014
Time on Study (years)
D
MRD < 0.1% (n = 28)
MRD 0.1-< 1% (n = 8)
MRD ≥ 1% (n = 37)
Probability
P = .0002
Time on Study (years)

New FHCRC policy

- If MFC negative: (a) no morphology to detect relapse, (b) morphology done only if declining counts
- Find MFC level above which morphology will invariably show >5% blasts by morphology (standard criterion for relapse) : eliminates need for morphology

Standardization

- FHCRC uses 10-color flow, Brent Wood is world expert etc.
- But morphology is not standardized either
  - discordance between pathologists in identifying blasts
  - disagreement on only 10/500 cells converts 4% blasts (no relapse) to 6% blasts (relapse)
  - 95% CI for 20/500 (4%)& 30/500 (6%) blasts overlap: 2-6% vs. 4-8%
  - MFC counts many more cells

MRD Questions

- Can its ability to improve prognostication be improved if done dynamically?
- Can it be used to assign treatment?
- Is its reduction useful?
- Can it replace morphology?
- Interrelations among different types

Inter-relationships Among Techniques to Detect MRD

	Days from initial Rx			
	21-28	29-35	36-42	> 42……
MFC				
Cyto				
FISH				
CGAT				
Molecular				

Recorded at each time as pos., neg., not done

Thanks

Drs. LoCoco and Sanz

Dr. Freireich

Dr. Kantarjian

Dr. Appelbaum