Optimal Pool Sizes for Group Testing

Christopher R. Bilder[1], Joshua M. Tebbs[2], Christopher S. McMahan[3], Brianna D. Hitt[1]
[1]University of Nebraska-Lincoln, [2]University of South Carolina, [3]Clemson University

Background

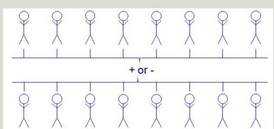
Abstract

Group testing is an indispensable tool for laboratories when testing high volumes of clinical specimens for infectious diseases. An important decision that needs to be made prior to its implementation is the group (pool) sizes to use. In best practice, an objective function is chosen and then minimized to determine an optimal set of group sizes. There are a few options for these objective functions, and they differ based on how the expected number of tests, assay characteristics, and laboratory constraints are taken into account. The purpose of this presentation is to closely examine a few common objective functions. We show the group sizes and/or results from using these objective functions are largely the same for standard testing algorithms in a wide variety of situations.

Corresponding author: Christopher R. Bilder, chris@chrisbilder.com, www.chrisbilder.com

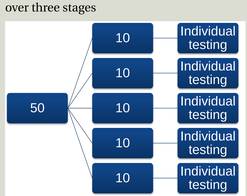
What is group testing?

- Used to screen a large number of individuals for an infectious disease
- Example: Blood donation screening for HIV, Hepatitis B, and Hepatitis C at the American Red Cross
  - An amalgamation of specimens from 16 individuals is a *group*
  - If this group tests negative, then all individuals are declared negative
  - If this group tests positive, then at least one individual is positive
    - Need to determine who is positive and who is negative
    - American Red Cross simply retests all group members individually
- Benefits in comparison to individual testing:
  - Smaller number of tests
  - Cost savings
- Need a small overall disease prevalence to prevent too many groups from testing positive

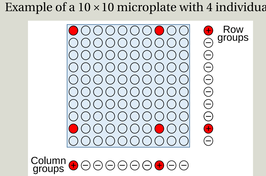


Algorithms

- Hierarchical
  - 2-stage: Dorfman (1943) testing used in the American Red Cross example
  - 3 or more stages are possible
  - Example: HIV testing in San Francisco (Sherlock et al. 2007) over three stages



- Array
  - Arrange specimens in a grid on a microplate
  - Test all specimens in one overall master group
  - If master group is positive, create groups by rows and columns and perform tests upon them
  - Intersections of positive testing rows and columns are retested individually
  - Example of a 10×10 microplate with 4 individual tests



Objective Functions

Purpose

- What group size(s) should be used for a testing configuration?
  - Want the smallest number of tests as possible ⇒ minimization of testing time and costs
  - Want the smallest number of testing errors as possible ⇒ maximization of accuracy
- Minimize an objective function to determine group size(s)
  - The resulting testing configuration is the *optimal testing configuration* (OTC)
  - Traditionally, the expected number of tests has been used
  - Recently, an alternative function was proposed by Malinovsky, Albert, and Roy (*Biometrics*, 2016)
  - This paper generated some controversy with replies written by
    - McMahan, Tebbs, and Bilder (*Biometrics*, 2016)
    - Hudgens (*Biometrics*, 2016)
  - All of these papers examined 2-stage hierarchical testing only
- Purpose: Compare the OTCs for different objective functions and commonly used group testing algorithms

Expected number of tests

- Define T as the total number of tests for an overall group of size I with a hierarchical algorithm
- OTC is found by minimizing the expected number of tests per individual: $O_{ET} = E(T)/I$
- Example: $E(T)$ for 3-stage hierarchical testing

$$E(T) = 1 + I_{11} \times P(G_{11} = 1) + \sum_{j=1}^{m_2} I_{2j} \times P(G_{11} = 1, G_{2j} = 1)$$

where
  - G_{sj} is the binary outcome (1 = positive, 0 = negative) for group j at stage s
  - I_{sj} is the size of group j at stage s
  - m_s is the number of groups at stage s
  - $P(G_{11} = 1)$ and $P(G_{11} = 1, G_{2j} = 1)$ are both functions of the number of groups, the overall disease prevalence p, and the assay sensitivity S_e and specificity S_p (details are given in Black, Bilder, and Tebbs, *JRSS-C*, 2015)
- Similar expressions can be obtained for $E(T)$ with a different number of stages or for array testing algorithms

Expected number of tests and correct classifications

- What about the number of correct classifications (accuracy)?
  - When using O_{ET} alone, one usually separately examines measures of accuracy like
    - Pooling sensitivity: $PS_e = P$(individual classified as positive by algorithm | true positive)
    - Pooling specificity: $PS_p = P$(individual classified as negative by algorithm | true negative)
  - All of these accuracy measures are functions of the group sizes and overall disease prevalence
- Malinovsky, Albert, and Roy (*Biometrics*, 2016)
  - Directly include the number of correct classifications (C) in the objective function
  - Minimize $O_{MAR} = E(T)/E(C)$
- O_{ET} vs. O_{MAR}
  - $C \le I$, $E(C) \le I$
  - C is close to I for most realistic applications
  - $E(C) = I[PS_p(1-p) + PS_e p]$ for equal group sizes at each stage in a hierarchical algorithm
  - $O_{ET} \le O_{MAR}$ for the same I

Comparisons

OTC for $p = 0.01$

Algorithm	S_e, S_p	Objective function	OTC*	$E(T)/I$	PS_e	PS_p
2-stage hierarchical	0.99	O_{ET}	11-1	0.2035	0.9801	0.9990
		O_{MAR}	11-1	0.2035	0.9801	0.9990
	0.95	O_{ET}	11-1	0.2351	0.9025	0.9932
		O_{MAR}	11-1	0.2351	0.9025	0.9932
	0.90	O_{ET}	12-1	0.2742	0.8100	0.9816
		O_{MAR}	12-1	0.2742	0.8100	0.9816
3-stage hierarchical	0.99	O_{ET}	25-5-1	0.1354	0.9703	0.9996
		O_{MAR}	25-5-1	0.1354	0.9703	0.9996
	0.95	O_{ET}	24-6-1	0.1443	0.8574	0.9973
		O_{MAR}	24-6-1	0.1443	0.8574	0.9973
	0.90	O_{ET}	24-6-1	0.1562	0.7290	0.9938
		O_{MAR}	24-6-1	0.1562	0.7290	0.9938

Algorithm	S_e, S_p	Objective function	OTC*	$E(T)/I$	PS_e	PS_p
Array w/o master	0.99	O_{ET}	25-1	0.1378	0.9703	0.9995
		O_{MAR}	25-1	0.1378	0.9703	0.9995
	0.95	O_{ET}	25-1	0.1475	0.8575	0.9970
		O_{MAR}	24-1	0.1475	0.8575	0.9972
	0.90	O_{ET}	25-1	0.1611	0.7291	0.9926
		O_{MAR}	24-1	0.1611	0.7291	0.9930
Array w/ master	0.99	O_{ET}	625-25-1	0.1364	0.9606	0.9995
		O_{MAR}	625-25-1	0.1364	0.9606	0.9995
	0.95	O_{ET}	625-25-1	0.1402	0.8146	0.9972
		O_{MAR}	576-24-1	0.1402	0.8146	0.9974
	0.90	O_{ET}	625-25-1	0.1450	0.6562	0.9934
		O_{MAR}	576-24-1	0.1450	0.6562	0.9937

* Equally sized groups were optimal at each stage; thus, a "24-6-1" means that stage 1 had a group size of 24, stage 2 had four groups of size 6, and stage 3 had twenty-four groups of size 1

- The same OTC is found for most cases
- When differences in the OTC occur, there is very little difference in accuracy

OTC for other values of p

- Table to the right shows the frequency of different OTCs for p =0.005, 0.01, ..., 0.15 (30 different p's)
- When differences occur, they typically happen for
  - Very low p
  - Unusually large p for a group testing application
  - Lower S_e and S_p cases that are infrequent in group testing applications

  and the OTCs with their corresponding accuracies are quite similar
- Similar findings occur for informative group testing when each individual has a different probability of positivity

Algorithm	S_e, S_p	Frequency	Largest difference* $E(T)/I$	PS_e	PS_p
2-stage hierarchical	0.99	0	–	–	–
	0.95	3	0.0018	0.0000	0.0049
	0.90	4	0.0023	0.0000	0.0054
3-stage hierarchical	0.99	0	–	–	–
	0.95	1	0.0014	0.0000	0.0051
	0.90	5	0.0051	0.0000	0.0098
Array w/o master	0.99	0	–	–	–
	0.95	5	0.0010	0.0018	0.0026
	0.90	8	0.0028	0.0022	0.0054
Array w/ master	0.99	2	0.0005	0.0006	0.0008
	0.95	4	0.0012	0.0017	0.0026
	0.90	8	0.0015	0.0018	0.0051

*$E(T)/I$, PS_e, and PS_p are always less for O_{ET} than for O_{MAR}

Conclusions

- Both objective functions result in the same or very similar OTCs
- O_{ET} may be preferred because
  - More meaningful for laboratories because they need to know $E(T)$ for planning purposes
  - Let A denote costs; then $E(A) \propto E(T)$ in many instances

This research was supported by Grant R01 AI121351 from the National Institutes of Health