The Optimal Group Size Controversy for Infectious Disease Testing: Much Ado About Nothing?!

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

Background

Abstract

Group testing, the process of testing specimen amalgamations, 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 involves determining what group sizes to use. In best practice, an objective function is chosen and then minimized to determine an optimal set of these group sizes, known as the optimal testing configuration (OTC). There are a few options for objective functions, and they differ based on how the expected number of tests, assay characteristics, and laboratory constraints are taken into account. These varied options have led to a recent controversy in the literature regarding which objective function is best. In our poster, we examine the most commonly proposed objective functions. We show that this controversy may be "much ado about nothing" because the OTCs, group sizes, and corresponding results (e.g., expected number of tests, accuracy measures) from using the two most commonly proposed objective functions are largely the same for standard testing algorithms in a wide variety of situations.

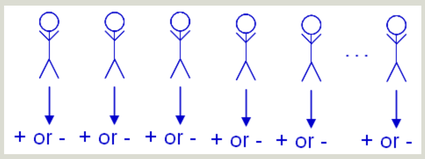
Corresponding author: Brianna D. Hitt, brianna.hitt@huskers.unl.edu

This research is supported in part by NIH Grant R01 AI121351.

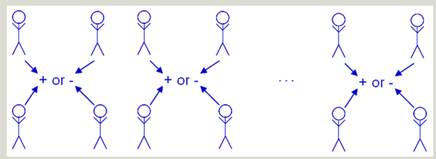
What is group testing?

- Used to screen a large number of individuals for an infectious disease or defect
- Practice of amalgamating specimens from individuals into *groups*
  - If this group tests negative, then all individuals are declared negative
  - If this group tests positive, then at least one individual is positive
- Saves time and resources in comparison to individual testing
- Human applications
  - Screening blood donations, detecting HIV treatment failure, testing for STDs, and surveilling for influenza
- Other applications
  - Testing for diseases in veterinary science, monitoring of West Nile virus in mosquitoes, detection of food contamination, and diagnosis of faulty network sensors

- Individual testing

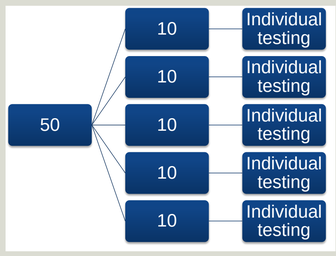


- Group (pooled) testing

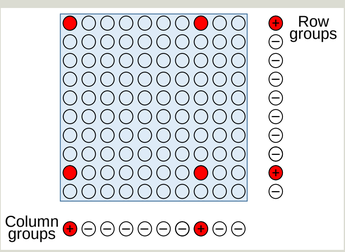


Algorithms

- Hierarchical
  - 2-stage: Dorfman (1943) testing used by the American Red Cross
  - 3 or more stages are possible
  - Example: HIV testing in San Francisco (Sherlock et al. 2007) using three stages
    - Initial group of 50 individuals
    - If group is positive, test 5 subgroups of 10 specimens each
    - If a subgroup is positive, test each of its individual members



- Array-based
  - Arrange specimens in a grid on a microplate
  - Amalgamate specimens by rows and columns and perform tests on them
  - Intersections of positive rows and columns are retested individually
    - Can test specimens in master pool before testing rows and columns
  - Example of a 10 x 10 microplate with 4 individual tests performed



Objective Functions

Purpose

- Choice of group sizes is important
  - Want the smallest number of tests possible ⇒ minimize testing time and costs
  - Want the smallest number of testing errors possible ⇒ maximize accuracy
- Group size(s) chosen by minimizing an objective function
  - Expected number of tests per individual is most common objective function
  - Resulting set of group sizes is the *optimal testing configuration* (OTC)
  - Alternative was recently proposed by Malinovsky, Albert, and Roy (*Biometrics*, 2016)
    - This paper generated controversy with replies by McMahan, Tebbs, and Bilder (2016) and Hudgens (2016) in *Biometrics*
  - All of these papers examined 2-stage hierarchical testing only
- Purpose: Compare the OTCs for different objective functions and commonly used group testing algorithms across a variety of situations

Expected number of tests

- Define T as the total number of tests needed to decode an overall group of size I
- 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, overall disease prevalence p, and assay sensitivity S_e and specificity S_p (Black et al., *JRSS-C*, 2015)
- Similar expressions can be obtained for $E(T)$ with different algorithms

Expected number of tests and correct classifications

- What about accuracy?
  - When using O_{ET}, one usually separately examines measures of accuracy like
    - Pooling sensitivity: $PS_e = P$(individual classified as positive | true positive)
    - Pooling specificity: $PS_p = P$(individual classified as negative | true negative)
    - Pooling positive and negative predictive values
  - 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}
  - $C \leq I$, $E(C) \leq I$
  - 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
  - $O_{ET} \leq O_{MAR}$ 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
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

Note: 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 shows frequency of different OTCs for $p = 0.005, 0.01, ..., 0.15$ (30 different probabilities)
- Differences most commonly occur 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

  and the OTCs with their corresponding accuracies are quite similar in these situations

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

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

- Similar findings occur for informative group testing when each individual has a different probability of positivity

Summary

R Functions

- Functions available in the `binGroup` package to reproduce our work
  - Find the OTC and calculate operating characteristics with O_{ET} and O_{MAR}
  - Available for standard group testing algorithms
  - Examples available at `www.chrisbilder.com/grouptesting`

Conclusions

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