

Editorial

Open Access

Advancements in Medical Statistics Promote Economies of Scale

Andrew Dellinger* and Herbert Pang

Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

Introduction

Modern medical studies, especially those intended for high impact journals, can easily cost millions of dollars (USD). For example, cost of drug development, including phase I to III trials, averages about USD 800 million [1,2]. Studies searching for single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) that contribute to disease are also costly. The full processing cost of an Affymetrix SNP 6.0 microarray, for example, can be \$600 [3] multiplied by the 6000 to 9000 patients typical for recent high impact publications of Genome-Wide Association Studies (GWAS) [4-9]. Still, researchers aim for even larger sample sizes because clear answers of variants' relationships with disease are not forthcoming given current sample sizes. For example, exome sequencing studies of 5300 and 6700 patients only yielded either very rare variants or variants with small effect size [10]. More traditional clinical studies, which measure clinical parameters to differentiate between groups of patients, may also have high costs due to recruitment, treatment, testing, and follow-up. They may also have high cost if a large sample size is required to achieve sufficient power for detection of significance in the questions of interest [11,12].

The burden of funding required for these studies strains the resources of even large pharmaceutical companies [4], wealthy corporations [13], and government funded groups, The Cancer Genome Atlas (TCGA) [14], the National Heart, Lung and Blood Institute and Exome Sequencing Project (ESP) [15], who host some of the largest current medical studies. To fund studies of increasing size, it is necessary to implement economies of scale. Economies of scale maintain or lower study cost while increasing the study's number of patients, number of potential analyses, and power. The majority of articles in this special issue increase economies of scale in one or more of these ways.

Recent Advances

Four studies in this issue increase their economy of scale by lowering the cost per patient. Two studies accomplished this by inexpensive data collection. Rogus et al. [16] retrieved data that was obtained from a health system in the normal course of business. In this way, over 600,000 measurements from over 39,000 patients were obtained at low additional cost. Such a large number of patients gave the study large power to detect the difference between case and control groups, which was found to be highly significant. Maetani and Gamel [12] used existing data from the Cancer Institute Hospital in Tokyo, which included surgery, follow-up, and survival of about 3600 patients for up to 50 years. The relatively low cost of obtaining existing data enabled them to obtain sufficient sample size to run analyses on three subgroups by age, 50 subgroups by demographic and prognostic factors, and on patients treated in different years with different regimens. Lifelong follow-up data was also cheaply and readily available, as required for validation of their Boag model [17]. Another two studies lowered their cost per patient by creating new study designs. Matched case-control studies can have high cost due to issues like follow-up costs until the end of recruitment. Sugihara et al. [18] created a study design that avoids this cost. Their dynamic registration method draws from a pool of potential patients while balancing the two groups of patients for all prognostic factors. Erbas et al. [19] avoided the cost of control recruitment and follow-up by implementing a case-crossover design for their asthma study. Given typical recruitment difficulties, the larger the study, the more resources this design will save.

Advancements in medical statistics also increase economies of scale by increasing power and/or the number of feasible analyses without increasing patient size. Meta-analysis increases power with very little increase in cost by combining existing studies into a single dataset of larger patient size. Meta-analysis also makes additional analyses feasible, when in a single study these analyses either cannot be run or cannot reach statistical significance. Sandoval and Zarate [11] list these and six more ways that economy of scale is increased by meta-analysis.

Rogus et al. [16], Ketchum et al. [20], and Maetani and Gamel [12] increased the ability of their studies to achieve significant results without increasing cost. The advancement of Rogus et al. [16] was to use Monte Carlo resampling to correct for intra-subject correlation. This correlation kept their highly negative z-statistic from being significant despite the large number of subjects in the study. Maetani and Gamel [12] developed the Boag model in such a way that the expense of long-term follow-up is not necessary to predict the overall survival curve and the long-term effects of the measured variables in a study. Ketchum et al. [20] developed a within-subject normal-mixture model, minimizing costs by making only a few subjects necessary to obtain significant results. This was achieved by incorporating data from multiple tests per subject into the model. The model has accurate estimation with as few as 20 patients. It uses all the data and can include all covariates.

Several studies in this issue increase the number of potential analyses without increasing cost. The study design of Erbas et al. [19] allows more analyses than a case-control design study of the same size. Particularly, they had more power to detect interactions of time-related risk factors with asthma. Halabi [21] developed an adjustment for type I error rate that preserves the rate at the selected alpha level through multiple intermediate analyses. The adjustment also allows flexibility in the allocation of the type I error rate. In this way, the cost of the study is maintained while maximizing the study's ability to detect meaningful changes. Also, the study becomes adaptable to intermediate discoveries without the need to make costly changes to the study or to design an entirely new study. Bergemann et al. [22] combined three test scores over four time points to make global normalized z-scores. These scores have more power, detecting significant differences at all time points when single test scores only detected significance in some time points. The global scores also yield lower p-values than in previous analyses.

*Corresponding author: Andrew Dellinger, Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA, E-mail: andrew.dellinger@duke.edu

Received January 18, 2013; Accepted January 21, 2013; Published January 26, 2013

Citation: Dellinger A, Pang H (2012) Advancements in Medical Statistics Promote Economies of Scale. J Biomet Biostat S7:e001. doi:10.4172/2155-6180.S7-e001

Copyright: © 2012 Dellinger A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

And they give the study better ability to describe longitudinal trends and determine the size of the difference between groups.

Discussion

The next major medical breakthroughs will likely be made in studies with many thousands of patients. To finance such studies and increase their potential for meaningful discoveries, they must have the study design and statistical methodology that maximize power and the ability to answer the question of interest. In other words, they must optimize economy of scale or these breakthroughs may be hindered by insufficient resources, such as money or patients. Medical statistics must continue to advance, as through the articles in this special issue, in order for major advancement to be made in most fields of medical research. It is encouraging that progress is being made on all fronts: increase in the detection of significance, increase in the sensitivity to pick up more accurate p-values, increase in the ability to perform more analyses with the same number of samples, and a decrease in the number of patients required to accomplish these advances. In conclusion, further advances in economies of scale must be made, especially in studies like GWAS, exome sequencing, and copy number detection, which search for rare biological occurrences.

References

- Pronker ES, Weenen TC, Commandeur HR, Osterhaus AD, Claassen HJ (2011) The gold industry standard for risk and cost of drug and vaccine development revisited. Vaccine 29: 5846-5849.
- Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D (2011) The cost of drug development: a systematic review. Health Policy 100: 4-17.
- 3. http://www.uthsc.edu/research/molecular_resource/
- Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.
- Uddin M, Sturge M, Rahman P, Woods MO (2011) Autosome-wide copy number variation association analysis for rheumatoid arthritis using the WTCCC highdensity SNP genotype data. J Rheumatol 38: 797-801.
- Jones AM, Howarth KM, Martin L, Gorman M, Mihai R, et al. (2012) Thyroid cancer susceptibility polymorphisms: confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24. J Med Genet 49: 158-163.
- Wokolorczyk D, Gliniewicz B, Sikorski A, Zlowocka E, Masojc B, et al. (2008) A range of cancers is associated with the rs6983267 marker on chromosome 8. Cancer Res 68: 9982-9986.

- Haiman CA, Le Marchand L, Yamamato J, Stram DO, Sheng X, et al. (2007) A common genetic risk factor for colorectal and prostate cancer. Nat Genet 39: 954-956.
- Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, et al. (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 39: 984-988.
- 10. Kaiser J (2012) Genetic Influences on Disease Remain Hidden. Science 338: 1016-1017.
- 11. Sandoval DC, Zarate CAT (2012) The Meta-Analysis in Medicine: Tools for Clinical and Experimental Research. J Biomet Biostat S7-011.
- 12. Maetani S, Gamel JW (2012) Parametric Cancer Survival Analysis Based on the Boag Model: A Clinician's View. J Biomet Biostat S7-017.
- Bodenheimer T (2000) Uneasy alliance: Clinical investigators and the pharmaceutical industry. N Engl J Med 342: 1539-1544.
- Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455: 1061-1068.
- 15. www.nhlbi.nih.gov/resources/exome.htm
- Rogus JJ, Lin S-F, Lacson Jr EK (2012) Efficient Monte Carlo Resampling of Continuous Data with a Dichotomous Treatment Variable. J Biomet Biostat S7-022.
- Boag JW (1949) maximum likelihood estimates of the proportion of patients cured by cancer therapy. J R Stat Soc Series B Stat Methodol 11: 15-53.
- Sugihara M, Morita S, Yamanouchi N, Sakai S, Ohba N, et al. (2012) Dynamic Registration Method with Balancing for Prognostic Factors in Observational Studies. J Biomet Biostat S7-012.
- Erbas B, Dharmage SC, O'Sullivan M, Akram M, Newbigin E, et al. (2012) A Case-Crossover Design to Examine the Role of Aeroallergens and Respiratory Viruses on Childhood Asthma Exacerbations Requiring Hospitalization: The Mapcah Study. J Biomet Biostat S7-018.
- Ketchum JM, Best AM, Ramakrishnan V (2012) A Within-Subject Normal-Mixture Model with Mixed-Effects for Analyzing Heart Rate Variability. J Biomet Biostat S7-013.
- Halabi S (2012) Adjustment on the Type I Error Rate for a Clinical Trial Monitoring for both Intermediate and Primary Endpoints. J Biomet Biostat S7-015.
- Bergemann TL, Bangirana P, Boivin MJ, Connett JE, Giordani BJ, et al. (2012) Statistical Approaches to Assess the Effects of Disease on Neurocognitive Function Over Time. J Biomet Biostat S7-016.

This article was originally published in a special issue, **Medical statistics: Clinical and experimental research** handled by Editor(s). Dr. Herbert Pang, Duke University, USA. Page 2 of 2