

n4Studies: Sample Size Calculation for an Epidemiological Study on a Smart Device

Chetta Ngamjarus, M.Sc.^{*,**}, Virasakdi Chongsuvivatwong, Ph.D.^{*}, Edward McNeil, M.Sc.^{*}

^{*} Epidemiology Unit, Faculty of Medicine, Prince of Songkla University, Songkhla 90110, ^{**}Department of Biostatistics and Demography, Khon Kaen University, Khon Kaen 40002, Thailand.

ABSTRACT

Objective: This study was to develop a sample size application (called “n4Studies”) for free use on iPhone and Android devices and to compare sample size functions between n4Studies with other applications and software.

Methods: Objective-C programming language was used to create the application for the iPhone OS (operating system) while JavaScript, jQuery mobile, PhoneGap and jstat were used to develop it for Android phones. Other sample size applications were searched from the Apple app and Google play stores. The applications’ characteristics and sample size functions were collected. Spearman’s rank correlation was used to investigate the relationship between number of sample size functions and price.

Results: “n4Studies” provides several functions for sample size and power calculations for various epidemiological study designs. It can be downloaded from the Apple App and Google play store. Comparing n4Studies with other applications, it covers several more types of epidemiological study designs, gives similar results for estimation of infinite/finite population mean and infinite/finite proportion from GRANMO, for comparing two independent means from BioStats, for comparing two independent proportions from EpiCal application. When using the same parameters, n4Studies gives similar results to STATA, epicalc package in R, PS, G*Power, and OpenEpi.

Conclusion: “n4Studies” can be an alternative tool for calculating the sample size. It may be useful to students, lecturers and researchers in conducting their research projects.

Keywords: Sample size, epidemiology, application, calculation, smartphone

Siriraj Med J 2016;68:160-170

E-journal: <http://www.sirirajmedj.com/ojs>

INTRODUCTION

Sample size calculation is an essential requirement for clinical and epidemiological research. An appropriate sample size enables the investigator (s) to detect a real effect with enough precision.¹⁻³ A small sample size may have a large standard error or inadequate power to detect an important difference while a large sam-

ple size may be unethical, especially in a clinical trial.⁴ Thus, sample size and the parameters used in its calculation need to be displayed in health research articles as requested by The Consolidated Standards of Reporting Trials (CONSORT) and The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE).^{5,6}

For many researchers, the formulas for sample size calculation are often too difficult to remember. Thus, most researchers prefer to use a software on a personal computer (PC). Sample size calculation is the most common question encountered by statisticians and epidemiologists in formal and non-formal consultations. The problem

Correspondence to: Chetta Ngamjarus
E-mail: nchett@kku.ac.th

Received 8 December 2015

Revised 12 February 2016

Accepted 16 March 2016

is often raised on various occasions when a PC and the necessary software are not available. A simple to use mobile application (app) for sample size calculation would thus be useful. Therefore, the aim of this study was to develop an application for calculating the sample size required for almost all epidemiological study designs on a smart device, and to compare the results with other mobile device applications (apps) and PC software that calculates sample size for epidemiological studies.

MATERIALS AND METHODS

Development of n4Studies

The formulas were obtained from standard textbooks and journal articles listed in the references. Objective-C programming language and Xcode software were used to create the application for the iPhone OS while JavaScript, jQuery mobile, PhoneGap, jstat and the Eclipse software were used to develop it for Android phones and tablets.⁷⁻¹¹

Features and parameters for comparison process

For comparison with other sample size applications, the Apple App store and the Google play store were searched using the search term: "sample size". Only applications written in the English language were included. Table 1 lists the types of studies and arbitrary parameter values used in the calculation of the sample size. Those parameter values were arbitrarily set for comparing results between applications. Statistical software for PC were also used to compare the results. Alpha (α) was set as 0.05 and beta (β) at 0.20.

Data collection and analysis

The information about the application's characteristics was collected from the Apple app and Google play stores. Sample size and other functions from sample size applications were investigated after installing on smart devices. Frequencies with percentage were used to describe categorical data and median with minimum (min.) and maximum (max.) were used to present continuous data. Spearman's rank correlation (r_s) was used to investigate the relationship between number of sample size functions and price. The

TABLE 1. Type of sample size calculation and values given for essential parameters.

Type of study	Objective of study	Parameter values
Descriptive	Estimate an infinite population mean ¹²	Standard deviation (SD) = 20, maximum tolerated error (delta) = 2
	Estimate a finite population mean ¹²	SD = 20, maximum tolerated error = 2, population size = 2,000
	Estimate an infinite a population proportion ¹²	Proportion = 0.4, maximum tolerated error = 0.04
	Estimate a finite a population proportion ¹²	Proportion = 0.4, maximum tolerated error = 0.04, population size = 2,000
Comparative	Comparing two independent means ¹³	Mean of group 1 = 130, SD of group 1 = 15, mean of group 2 = 120, SD of group 2 = 20
	Comparing two dependent means ¹⁴	SD = 20, difference in means between two groups = 5
	Comparing two independent proportions (with continuity correction) ¹³	Proportion of group 1 = 0.15, proportion of group 2 = 0.05
	Comparing two independent proportions (without continuity correction) ¹⁵	Proportion of group 1 = 0.15, proportion of group 2 = 0.05
Matched case-control ¹⁶	Comparing proportion of exposure between case and control group	Odds ratio = 2, proportion of exposure = 0.15, ratio between case and control = 1

epicalc package on the R language and environment were used for all analyses.^{17,18}

RESULTS

n4Studies features

“n4Studies” requires iOS 5.0 or Android version 2.2 or later. Its screens were designed to have four tab menus at the bottom of the screen as shown in Figure 1. The “Sample size”, “Power of study” and “2×2 Table” tab open a menu for relevant calculations, whereas the tab “About us” gives other information related to the application

including questions and answers (Q&A) on the upper right corner (Fig 1-3). All calculation results are printed out in a textbox ready to send via email (Fig 4 and 5).

Comparison of features and results of calculation between n4Studies and other free sample size apps

On 28 April 2014, there were 84 applications (37 iOS applications and 47 android applications) identified from two majors app stores (Apple app and Google play stores). Seventeen iOS and 37 android applications were excluded

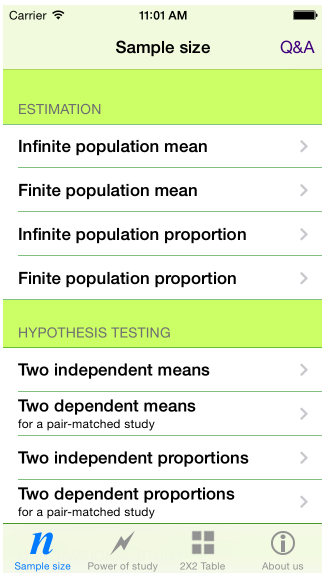


Fig 1. List of sample size calculations

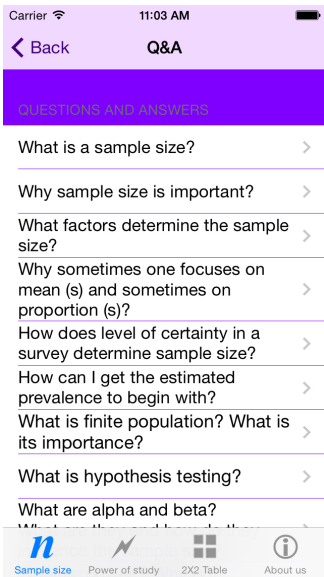


Fig 3. Q&A screen

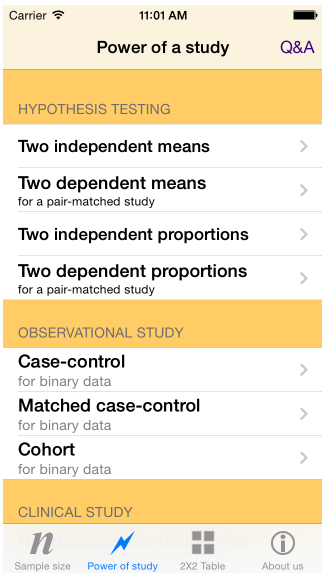


Fig 2. List of power calculations

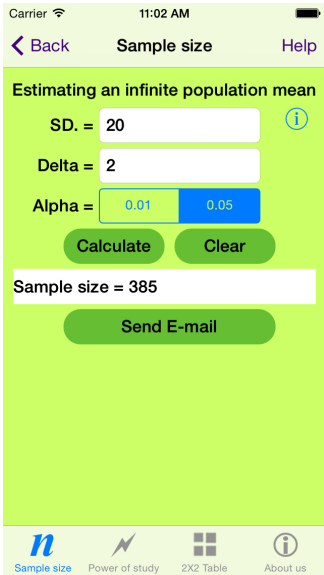


Fig 4. Sample size calculation screen

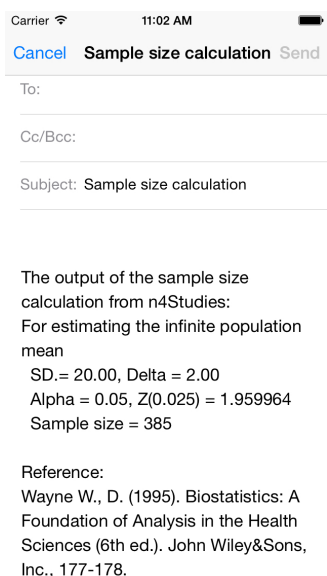


Fig 5. Screen for email a result

due to being irrelevant or not in English. After removal of duplications in two OS, 21 applications were examined (Fig 6).

Among these 21 applications, 10 were free: n4Studies, Sample size, N Sample size, SampSize, GLIMMPSE Lite, GiPA, WhatStat, STATCALC, Samclus and A+A Stats.¹⁹⁻²⁸ The remaining 11 were commercial (Sample size calculator App, 95/95 Sample size calculator, Sample size calculator for clinical research, EpiCal, My Sample Size, GRANMO, Power Analysis, StaStat, Biostats Calculator, Sample Size Calculator - S-o-m (SSize Calculator), Med calc 3000 EBM Stats).²⁹⁻³⁹ As shown in Table 2 the median minimum price of these 11 commercial applications was US \$ 1.99. The median number

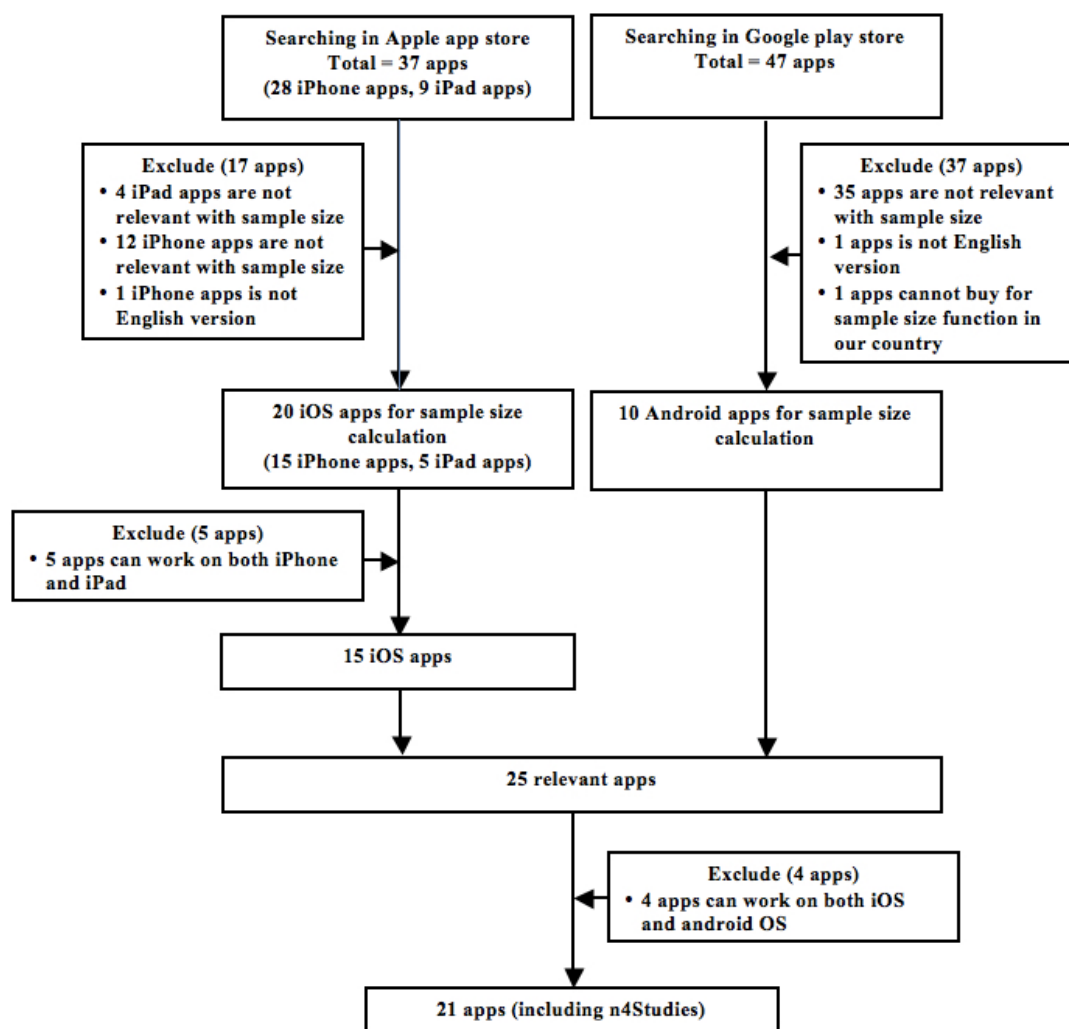


Fig 6. Flow of selecting the included apps

TABLE 2. Sample size apps characteristic (n (%))

Variable	Free App	Commercial App	Total
Total apps	10	11	21
Minimum price (\$)♣	-	1.99 (0.99, 9.99)	0.99 (9.99)
OS			
iOS	3 (30)	8 (72.7)	11 (52.4)
Android	4 (40)	2 (18.2)	6 (28.2)
Both iOS and android	3 (30)	1 (9.1)	4 (19.0)
OS version requirement♣			
iOS version	5.0 (4.3, 6.1)	5.0 (3.0, 7.0)	5.0 (3.0, 7.0)
Android version	1.9 (1.0, 4.0)	1.5 (1.1, 2.2)	1.6 (1.0, 4.0)
No. of sample size calculation function♣			
Without n4Studies	1 (1, 10)	3 (1, 17)	2 (1, 17)
With n4Studies	1.5 (1, 17)	3 (1, 17)	2 (1, 17)
Other statistical function			
No	5 (50)	4 (36.4)	9 (42.9)
Yes	5 (50)	7 (63.6)	12 (57.1)
App size (MB.)♣	1.35 (0.045, 2.3)	1.3 (0.151, 2.7)	1.3 (0.05, 2.7)
Provide reference in each formula			
No	7 (70)	10 (90.9)	17 (81)
Yes	3 (30)	1 (9.1)	4 (19)
Save output			
No	8 (80)	10 (90.9)	18 (85.7)
Yes	2 (20)	1 (9.1)	3 (14.3)
Email output			
No	6 (60)	9 (81.8)	15 (71.4)
Yes	4 (40)	2 (18.2)	6 (28.6)
Provide help or tutorial or instruction			
No	3 (30)	4 (36.4)	7 (33.3)
Yes	7 (70)	7 (63.6)	14 (66.7)
Provide other content in app (such as Q&A)			
No	9 (90)	11 (100)	20 (95.2)
Yes	1 (10)	-	1 (4.8)
Launch time (year)			
2009	0 (0)	2 (20)	2 (11.8)
2010	1 (14.3)	0 (0)	1 (5.9)
2011	1 (14.3)	3 (30)	4 (23.5)
2012	2 (28.5)	3 (30)	5 (29.4)
2013	3 (42.9)	2 (20)	5 (29.4)
Last update (year)			
2010	2 (20)	1 (9.1)	3 (14.3)
2011	1 (10)	3 (27.3)	4 (19)
2012	1 (10)	1 (9.1)	2 (9.5)
2013	2 (20)	5 (45.5)	7 (33.3)
2014	4 (40)	1 (9.1)	5 (23.8)

Note: ♣ Present as Median (min., max.)

of sample size calculation functions on all applications was 2 (range 1-17) and 12 (57.1%) had other functions. The median application size was 1.3 megabytes (MB.) (range 0.05-2.7 MB.). Four apps provided a reference for each sample size calculation, three apps could save the output, six apps could email its output, 14 apps had help pages and one app (n4Studies) provided questions and answers for the users. More than 50% of all applications were first published after 2011 and have been updated after 2012 (Table 2). Furthermore, there was a small positive relationship between number of sample size functions and price ($r_s = 0.127$) (Fig 7).

For the sample size calculation feature, almost half of the applications could be used for calculating the sample size required for compare two independent means. Except for n4Studies, SampSize, WhatStat, Sample size calculator for clinical research, EpiCal, and GRANMO, all have only few functions (Table 3). Based on the arbitrary parameters shown in Table 1, n4Studies, N Sample size, EpiCal, GRANMO and Power

Analysis gave the same sample size for estimating an infinite population mean, and n4Studies and GRANMO gave a similar result for estimating a finite population mean. Sample size calculation app, EpiCal and GRANMO gave similar sample sizes to n4Studies for estimating an infinite population proportion. Sample size calculation app, GRANMO and n4Studies gave the same sample size for estimating a finite population proportion. BioStats Calculator and n4Studies gave the same sample size for comparing two independent means. EpiCal gave a similar sample size to n4Studies for comparing two independent proportions (as marked by an asterisk (*) in Table 3). Moreover, all other applications required different parameters, and thus they were not comparable. For free applications, SampSize and n4Studies have more sample size features than the others. Among the commercial applications, Sample size calculator for clinical research, EpiCal and GRANMO have more sample size functions than the others (Table 3).

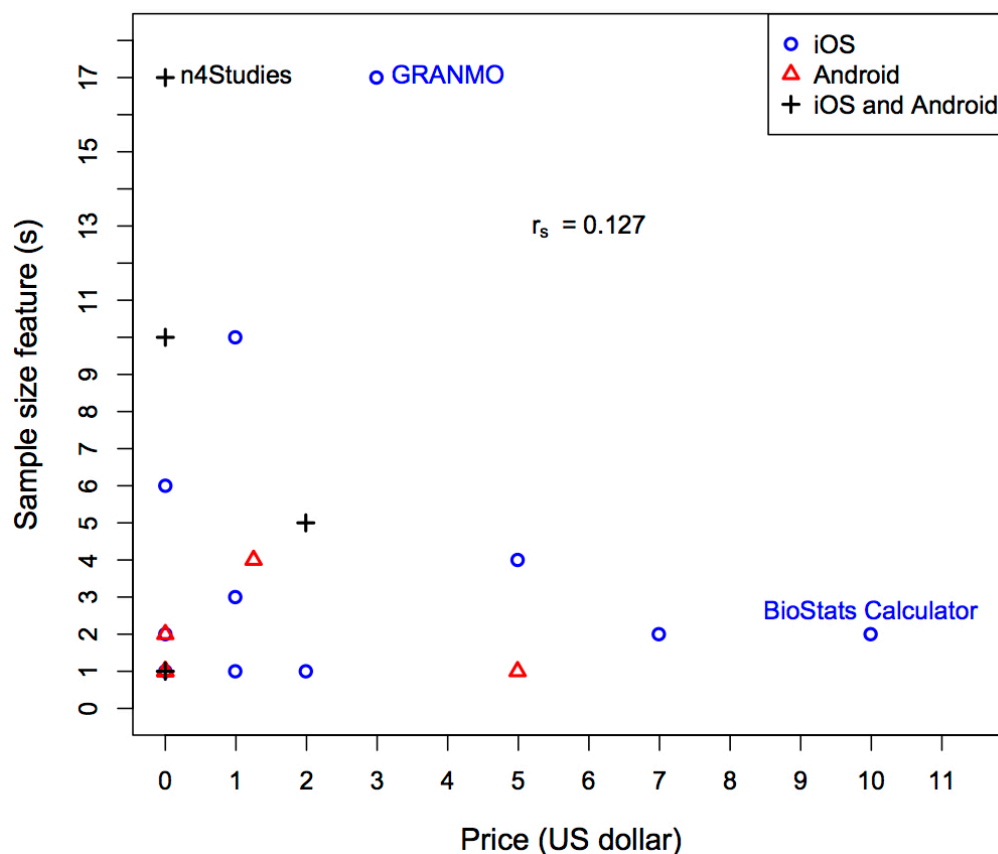


Fig 7. Relationship between number of sample size functions and price

TABLE 3. Comparisons of features available on the mobile applications[illegible]

TABLE 3. Comparisons of features available on the mobile applications

Formula	Free app											Commercial app										
	n4Studies	Sample size	N Sample size	SampleSize	GLIMMPSE Lite	G*PA	WhatStat	STATCALC	SamPlus	A+A Stats	Sample size calculator app	95/95 Sample size calculator	Sample size calculator for clinical	EpiCal	My Sample Size	GRANMO	Power Analysis	StaStat	BioStats Calculator	Size Calculator	Med calc 3000 EBM Stats	
Cluster randomized trial	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	✓	-	-	-	-	-	
Survival	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	✓	-	✓	-	-	-	
Bioequivalence	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
For size of random sampling error	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Based on margin of error	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	
For detecting disease	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	
Hypothesis testing for one population mean	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	✓	-	-	-	✓	-	
Hypothesis testing for one population proportion	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	✓	-	-	-	✓	-	
For correlation coefficient	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	
Repeated measurements	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-	
Based on confidence interval	-	✓	-	-	-	-	-	✓	-	✓	✓	✓	-	-	✓	-	-	-	-	-	✓	
Other feature																						
Provide reference in each formula	✓	-	-	✓	-	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	✓	-	
Provide other statistical function																						
Power of a study	✓	-	-	✓	✓	-	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	
Statistic from 2 by 2 table	✓	-	-	-	-	-	✓	-	-	-	-	-	-	✓	-	-	-	✓	-	-	✓	
Other (e.g. hypothesis testing, correlation)	-	-	-	-	-	-	-	✓	-	-	-	-	-	-	-	-	✓	✓	✓	-	✓	
Provide help or tutorial or instruction	✓	✓	-	✓	✓	-	-	✓	✓	✓	✓	-	✓	✓	✓	-	✓	✓	✓	-	-	
Provide other content in app (such as Q&A)	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Save output	-	-	-	-	✓	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	
Email output	✓	-	-	✓	✓	-	✓	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	
Provide information for contacting author (s)	✓	✓	-	-	✓	-	✓	✓	-	✓	-	-	-	✓	-	-	✓	-	-	-	-	

Note: ✓ Function is available on the app, * Giving result equal or close to that from n4Studies.

TABLE 4. Comparing result between n4Studies and statistical software

Formula	Free software n4Studies ¹⁹	Epicalc on R ¹⁸	PS ⁴⁰	G*Power ⁴¹	Commercial software OpenEpi ⁴²	STATA ⁴³
Estimating an infinite population mean	385	-	-	-	-	-
Estimating a finite population mean	323	-	-	-	-	-
Estimating an infinite a population proportion	577	576	-	-	576	-
Estimating a finite a population proportion	448	447	-	-	448	-
Comparing two independent means	n1 = 50, n2 = 50	n1 = 50, n2 = 50	DP	n1 = 51, n2 = 51	n1 = 50, n2 = 50	n1 = 50, n2 = 50
Comparing two dependent means	126	-	128	128	-	126
Comparing two independent proportions (with continuity correction) ^ψ	n1 = 141, n2 = 141	-	n1 = 140, n2 = 140	n1 = 141, n2 = 141	n1 = 141, n2 = 141	n1 = 141, n2 = 141
Comparing two independent proportions (with continuity correction) ^ψ	n1 = 160, n2 = 160	n1 = 160, n2 = 160	n1 = 159, n2 = 159 (Fisher's exact)	n1 = 151, n2 = 151 (Fisher's exact)	n1 = 160, n2 = 160	n1 = 160, n2 = 160
Matched case-control study	206	-	-	-	-	-

Note: DP: the software requested different parameters from table 1, ψ is applicable to case-control study, cohort study and RCT

Comparing results between n4Studies and other statistical software

n4Studies, epicalc and EpiOpen gave a similar sample sizes required when estimating an infinite and finite population proportion. For comparing two independent means, most software gave the same sample size except for PS because it requires different parameters for setting as shown in Table 1 and thus it's results could not be compared with other software. PS and G*Power calculate sample size for comparing two independent proportions based on Fisher's exact test so it gives smaller sample sizes than the other software. All software could estimate the sample size required for a case-control study, cohort study and randomized controlled trial (RCT) since the formulas are the same as for comparing two independent proportions (Table 4).

DISCUSSION

This study revealed that during the review period most available free applications are not adequate for clinical epidemiological sample size calculation except n4Studies, SampSize and WhatStat. n4Studies gave the same results as those from standard sample size calculation software on the PC or using applications available on the internet (OpenEpi). For a serious statistician or epidemiologist, often more than one PC software is used. A mobile application can be a supplement for those who are outside the office, or when the internet is not accessible. For a less experienced person, n4Studies is easy to download, install and use. It can be an alternative tool for learning. Moreover, the application's price is not related to the number of sample size functions. Some free applications including n4Studies have more sample size functions than commercial applications. Most of the free sample size applications do not need a high performance of smart device because they only need a small memory and not a high OS version, especially for android devices. In addition, only four applications (n4Studies, SampSize, Samclus, and Ssize calculator) provided references for each formula.

The special strength of n4Studies includes covering a broader range of study designs, avail-

ability of formulas and references, Q&A and direct e-mail interface. The limitations are lacking of tutorial and examples. This app is, therefore, most suitable for those who already have a certain theoretical background in sample size calculation. This study was done in 2014 after the smartphone and tablet started to be popular. In the future, with higher computing capacity, it is expected that there will be more features available for researchers to choose and utilize.

CONCLUSION

In conclusion, most current applications for calculation of sample size have limited capability compared to those on a PC. The exception is n4Studies, a mobile application which can meet most of the requirements met on a PC and gives practically the same results. n4Studies can be an alternative tool for calculating the sample size. It is friendly and easy to use and takes only a minute to run. It may be useful to students, lecturers and researchers in conducting their research projects.

ACKNOWLEDGEMENTS

We would like to thank The Royal Golden Jubilee Ph.D. Program - The Thailand Research Fund for providing the funding to conduct this project. We thank Dr. Alan Geater and Mr. Tagoon Prappre for their valuable suggestions during the development of this application. Epidemiology Unit is partially supported by the National Science and Technology Development Agency through the Research Chair Grant provided to Professor Virasakdi Chongsuvivatwong.

REFERENCES

- Wittes J. Sample size calculations for randomized controlled trials. *Epidemiol Rev.* 2002;24(1):39-53.
- Sathian B, Sreedharan J, Baboo SN, Sharan K, Abhilash ES, Rajesh E. Relevance of sample size determination in medical research. *Nepal Journal of Epidemiology.* 2010; 1(1).
- Lan L, Lian Z. Application of statistical power analysis - How to determine the right sample size in human health, comfort and productivity research. *Building and Environment.* 2010;45:1202-13.
- Altman DG. Statistics and ethics in medical research: III How large a sample? *British Medical Journal.* 1980;281: 1336-8.
- Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Medical Research Methodology.* 2001;1(1).
- Elm Ev, Altman DG, Egger M, Pocock SJ, GÄtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *BMJ* 2007; 335:806-8.
- Xcode: Apple's integrated development environment [computer program]. Version 5.0.2: Apple Inc.; 2013.
- PhoneGap [computer program]. Adobe Systems Inc.; 2014.
- jQuery mobile [computer program]. The jQuery Foundation; 2014.
- jStat [computer program]. <http://www.jstat.org>; 2014.
- Eclipse: Android developer tools [computer program]. The Android Open Source Project; 2014.
- Daniel WW. Biostatistics: A foundation for analysis in the health sciences. ohn Wiley&Sons, Inc; 1995.
- Rosner B. Fundamentals of biostatistics. Duxbury: Thomson learning; 2000.
- Chow S-C, Shao J, Wang H. Sample size calculations in clinical research. Chapman & Hall/CRC; 2003.
- Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. John Wiley&Sons; 2003.
- Parker RA, Bregman DJ. Sample size for individually matched case-control studies. *Biometrics* 1986;42:919-26.
- R: A language and environment for statistical computing [computer program]. R Core Team; 2014.
- epicalc: Epidemiological calculator [computer program]. 2012.
- n4Studies: iPhone application [computer program]. Version 1.1.0: The Royal Golden Jubilee Ph.D. Program - The Thailand Research Fund and Prince of Songkla University; 2014.
- Sample size: iPhone application [computer program]. Version 1.0: Stats Assistance; 2012.
- N: iPhone application [computer program]. Version 1.1.2012.
- SampSize: iPhone application [computer program]. Version 1.1.4: Epi Genesys Limited; 2013.
- GLIMMPSE Lite: iPhone application [computer program]. Version 1.0.02013.
- WhatStat: iPhone application [computer program]. Version 1.1: everywhereIM Ltd.; 2014.
- GiPA: Sample size calculator [computer program]. ENYEL; 2011.
- STATCALC: Statistical calculator [computer program]. Version 1.2: Market-i; 2014.
- Samclus: Android application [computer program]. Andstat; 2013.

28. A+A Stats: Android application [computer program]. A+A; 2014.
29. Sample size calculator App: iPhone application [computer program]. Version 2.5: LHC Apps; 2013.
30. 95/95: iPhone application [computer program]. Version 1.0: Helios Apps; 2011.
31. Sample size calculator for clinical research: iPhone application [computer program]. Version 1.0: Manon Girard; 2011.
32. EpiCal: iPhone application [computer program]. Version 1.0: Michael Reichel; 2012.
33. GRANMO: iPhone application [computer program]. Version 1.2.5: Can Antaviana; 2014.
34. My Sample Size: iPhone application [computer program]. Version 1.0: LulaSoft; 2013.
35. Power Analysis: iPhone application [computer program]. Version 1.4.0: A&X Analysis, LLC; 2013.
36. StaStat: iPhone application [computer program]. Version 2.7: Yasunari Miyagi; 2013.
37. Biostats Calculator: iPhone application [computer program]. Version 1.3.1: Sam McCall; 2010.
38. Sample Size Calculator - S-o-m: android Application [computer program]. Version 1.0: Stat-o-matic; 2013.
39. Med calc 3000 EBM Stats: Android application [computer program]. Version 10.3: Foundation Internet Services; 2013.
40. PS: Power and sample size calculation [computer program]. Version 3.1.22014.
41. Faul F, Erdfelder E, Lang A.-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*. 2007;39:175-191.
42. Dean AG, Sullivan KM, MM S. OpenEpi: Open source epidemiologic statistics for public health. 2014; www.openepi.com, 2014.
43. Stata statistical software [computer program]. Version 10: College Station, TX: StataCorp LP; 2007.