SAMPLE SIZE DETERMINATION IN HEALTH STUDIES

A Practical Manual

S. K. Lwanga

Epidemiological and Statistical Methodology
World Health Organization
Geneva, Switzerland

and

S. Lemeshow

Division of Public Health
University of Massachusetts at Amherst
MA, USA

World Health Organization
Geneva
1991

WHO Library Cataloguing in Publication Data

Lwanga, S. K.
Sample size determination in health studies : a practical manual.

1.Sampling studies 2.Health surveys I.Lemeshow, S. II.Title

ISBN 92 4 154405 8 (NLM Classification: WA 950)

Typeset in India
Printed in England
89/8087–Macmillan/Clays–5000

Contents

Preface

In many of WHO's Member States, surveys are being undertaken to obtain information for planning, operating, monitoring and evaluating health services. Central to the planning of any such survey is the decision on how large a sample to select from the population under study, and it is to meet the needs of health workers and managers responsible for making that decision that this manual has been prepared. It is essentially a revised and expanded version of a popular unpublished document on sample size that has been widely used in WHO's field projects and training courses. The examples and tables presented, which have been selected to cover many of the approaches likely to be adopted in health studies, will not only be of immediate practical use to health workers but also provide insight into the statistical methodology of sample size determination.

The authors would like to thank Dr B. Grab, formerly Statistician, WHO, Geneva, Dr R. J. Hayes, London School of Hygiene and Tropical Medicine, and colleagues in the Unit of Epidemiological and Statistical Methodology, the Diarrhoeal Diseases Control Programme and the Expanded Programme on Immunization of WHO for their comments. The financial support of the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases is gratefully acknowledged.

Introduction

Among the questions that a health worker should ask when planning a survey or study is "How large a sample do I need?" The answer will depend on the aims, nature and scope of the study and on the expected result, all of which should be carefully considered at the planning stage.

For example, in a study of the curative effect of a drug on a fatal disease such as the acquired immunodeficiency syndrome (AIDS), where a single positive result could be important, sample size might be considered irrelevant. In contrast, if a new malaria vaccine is to be tested, the number of subjects studied will have to be sufficiently large to permit comparison of the vaccine's effects with those of existing preventive measures.

The type of "outcome" under study should also be taken into account. There are three possible categories of outcome. The first is the simple case where two alternatives exist: yes/no, dead/alive, vaccinated/not vaccinated, existence of a health committee/lack of a health committee. The second category covers multiple, mutually exclusive alternatives such as religious beliefs or blood groups. For these two categories of outcome the data are generally expressed as percentages or rates. The third category covers continuous response variables such as weight, height, age and blood pressure, for which numerical measurements are usually made. In this case the data are summarized in the form of means and variances or their derivatives. The statistical methods appropriate for sample size determination will depend on which of these types of outcome the investigator is interested in.

Only once a proposed study and its objectives have been clearly defined can a health worker decide how large a sample to select from the population in question. This manual is intended to be a practical guide to making such decisions. It presents a variety of situations in which sample size must be determined, including studies of population proportion, odds ratio, relative risk and incidence rate.[1] In each case the information needed is specified and at least one illustrative example is given. All but one example are accompanied by tables of minimum sample size for various study conditions so that the reader may obtain solutions to problems of sample size without recourse to calculations (more extensive tables are available in the publication by Lemeshow et al. mentioned below). Random sampling is assumed for all examples, so that if the sample is not to be selected in a statistically random manner the tables are not valid.

[1] Continuous response variables are not considered in this manual because of the wide range of possible parameter values.

The manual is designed to be used in "cookbook" fashion; it neither helps the reader to decide what type of study, confidence level or degree of precision is most appropriate, nor discusses the theoretical basis of sample size determination. Before using the manual, therefore, the investigator should have decided on the study design, made a reasonable guess at the likely result, determined what levels of significance, power and precision (where relevant) are required and considered operational constraints such as restrictions on time or resources. The reader who wishes to learn more about the statistical methodology of sample size determination is referred to Lemeshow, S. et al., *Adequacy of sample size in health studies* (Chichester, John Wiley, 1990; published on behalf of the World Health Organization) or to any standard textbook on statistics.

One-sample situations

Estimating a population proportion with specified absolute precision

Required information and notation

(*a*) Anticipated population proportion	P
(*b*) Confidence level	$100(1-\alpha)\%$
(*c*) Absolute precision required on either side of the proportion (in percentage points)	d

A rough estimate of P will usually suffice. If it is not possible to estimate P, a figure of 0.5 should be used (as in Example 2); this is the "safest" choice for the population proportion since the sample size required is largest when $P=0.5$. If the anticipated proportion is given as a range, the value closest to 0.5 should be used.

Tables 1a and 1b (pages 25–26) present minimum sample sizes for confidence levels of 95% and 90%, respectively.

Simple random sampling is unlikely to be the sampling method of choice in an actual field survey. If another sampling method is used, a larger sample size is likely to be needed because of the "design effect". For example, for a cluster sampling strategy the design effect might be estimated as 2. This would mean that, to obtain the same precision, twice as many individuals would have to be studied as with the simple random sampling strategy. In Example 2, for instance, a sample size of 192 would be needed.

Example 1

A local health department wishes to estimate the prevalence of tuberculosis among children under five years of age in its locality. How many children should be included in the sample so that the prevalence may be estimated to within 5 percentage points of the true value with 95% confidence, if it is known that the true rate is unlikely to exceed 20%?

Solution

(*a*) Anticipated population proportion	20%
(*b*) Confidence level	95%
(*c*) Absolute precision (15%–25%)	5 percentage points

Table 1a (page 25) shows that for $P=0.20$ and $d=0.05$ a sample size of 246 would be needed.

If it is impractical, with respect to time and money, to study 246 children, the investigators should lower their requirements of confidence to, per-

haps, 90%. Table 1b (page 26) shows that, in this case, the required sample size would be reduced to 173.

Example 2

An investigator working for a national programme of immunization seeks to estimate the proportion of children in the country who are receiving appropriate childhood vaccinations. How many children must be studied if the resulting estimate is to fall within 10 percentage points of the true proportion with 95% confidence? (It is not possible to make any assumption regarding the vaccination coverage.)

Solution

(*a*) Anticipated population proportion ("safest" choice, since P is unknown)	50%
(*b*) Confidence level	95%
(*c*) Absolute precision (40%–60%)	10 percentage points

Table 1a (page 25) shows that for $P=0.50$ and $d=0.10$ a sample size of 96 would be required.

Estimating a population proportion with specified relative precision

Required information and notation

(*a*) Anticipated population proportion	P
(*b*) Confidence level	$100(1-\alpha)\%$
(*c*) Relative precision	ε

The choice of P for the sample size computation should be as "conservative" (small) as possible, since the smaller P is the greater is the minimum sample size.

Tables 2a and 2b (pages 27–28) present minimum sample sizes for confidence levels of 95% and 90%, respectively.

Example 3

An investigator working for a national programme of immunization seeks to estimate the proportion of children in the country who are receiving appropriate childhood vaccinations. How many children must be studied if the resulting estimate is to fall within 10% (not 10 percentage points) of the true proportion with 95% confidence? (The vaccination coverage is not expected to be below 50%.)

Solution

(*a*) Anticipated population proportion (conservative choice)	50%
(*b*) Confidence level	95%
(*c*) Relative precision (from 45% to 55%)	10% (of 50%)

Table 2a (page 27) shows that for $P=0.50$ and $\varepsilon=0.10$ a sample size of 384 would be needed.

If it is impractical, with respect to time and money, to study 384 children, the investigators should lower their requirements of confidence to, per-

haps, 90%. Table 2b (page 28) shows that, in this case, the required sample size would be reduced to 271.

Simple random sampling is unlikely to be the sampling method of choice in an actual field survey. If another sampling method is used, a larger sample size is likely to be needed because of the "design effect". For example, for a cluster sampling strategy the design effect might be estimated as 2. This would mean that, to obtain the same precision, twice as many individuals would have to be studied as with the simple random sampling strategy. In this example, therefore, for a confidence level of 95%, a sample size of 768 would be needed.

Example 4

How large a sample would be required to estimate the proportion of pregnant women in a population who seek prenatal care within the first trimester of pregnancy, to within 5% of the true value with 95% confidence? It is estimated that the proportion of women seeking such care will be between 25% and 40%.

Solution

(*a*) Anticipated population proportion	25%–40%
(*b*) Confidence level	95%
(*c*) Relative precision	5% (of 25%–40%)

Table 2a (page 27) presents the following sample sizes for $\varepsilon = 0.05$ and for population proportions in the range 25%–40%.

P	Sample size
0.25	4610
0.30	3585
0.35	2854
0.40	2305

Therefore a study of roughly 4610 women might be planned to satisfy the stated objectives. If necessary a smaller sample size could be used, but this would result in a loss of precision or confidence or both if the true value of P was close to 25%.

Hypothesis tests for a population proportion

This section applies to studies designed to test the hypothesis that the proportion of individuals in a population possessing a given characteristic is equal to a particular value.

Required information and notation

(*a*) Test value of the population proportion under the null hypothesis	P_0
(*b*) Anticipated value of the population proportion	P_a
(*c*) Level of significance	$100\alpha\%$
(*d*) Power of the test	$100(1-\beta)\%$
(*e*) Alternative hypothesis: either	$P_a > P_0$ or $P_a < P_0$ (for one-sided test)
or	$P_a \neq P_0$ (for two-sided test)

Tables 3a–d (pages 29–32) present minimum sample sizes for a level of significance of 5%, powers of 90% and 80%, and both one-sided and two-sided tests. For Tables 3c and 3d the complement of P_0 should be used as the column value whenever $P_0 > 0.5$.

Example 5

The five-year cure rate for a particular cancer (the proportion of patients free of cancer five years after treatment) is reported in the literature to be 50%. An investigator wishes to test the hypothesis that this cure rate applies in a certain local health district. What minimum sample size would be needed if the investigator was interested in rejecting the null hypothesis only if the true rate was less than 50%, and wanted to be 90% sure of detecting a true rate of 40% at the 5% level of significance?

Solution

(*a*) Test cure rate	50%
(*b*) Anticipated cure rate	40%
(*c*) Level of significance	5%
(*d*) Power of the test	90%
(*e*) Alternative hypothesis (one-sided test)	cure rate < 50%

Table 3a (page 29) shows that for $P_0 = 0.50$ and $P_a = 0.40$ a sample size of 211 would be needed.

Example 6

Previous surveys have demonstrated that the usual prevalence of dental caries among schoolchildren in a particular community is about 25%. How many children should be included in a new survey designed to test for a decrease in the prevalence of dental caries, if it is desired to be 90% sure of detecting a rate of 20% at the 5% level of significance?

Solution

(*a*) Test caries rate	25%
(*b*) Anticipated caries rate	20%
(*c*) Level of significance	5%
(*d*) Power of the test	90%
(*e*) Alternative hypothesis (one-sided test)	caries rate < 25%

Table 3a (page 29) shows that for $P_0 = 0.25$ and $P_a = 0.20$ a sample size of 601 would be needed.

If the investigators use this sample size, and if the actual caries rate is less than 20%, then the power of the test will be larger than 90%, i.e. they will be more than 90% likely to detect that rate.

Example 7

The success rate for a surgical treatment of a particular heart condition is widely reported in the literature to be 70%. A new medical treatment has been proposed that is alleged to offer equivalent treatment success. A hospital without the necessary facilities or staff to provide the surgical treatment has decided to use the new medical treatment for all new patients presenting with this condition. How many patients must be studied to test the hypothesis that the success rate of the new method of treatment is 70% against an alternative hypothesis that it is not 70% at the 5% level of significance? The investigators wish to have a 90% power of detecting a difference between the success rates of 10 percentage points or more in either direction.

Solution

(*a*) Test success rate	70%
(*b*) Anticipated success rate	80% or 60%

(*c*) Level of significance	5%
(*d*) Power of the test	90%
(*e*) Alternative hypothesis (two-sided test)	success rate $\neq 70\%$

Table 3c (page 31) shows that for $(1-P_0)=0.30$ and $|P_a-P_0|=0.10$ a sample size of 233 would be needed.

Example 8

In a particular province the proportion of pregnant women provided with prenatal care in the first trimester of pregnancy is estimated to be 40% by the provincial department of health. Health officials in another province are interested in comparing their success at providing prenatal care with these figures. How many women should be sampled to test the hypothesis that the coverage rate in the second province is 40% against the alternative that it is not 40%? The investigators wish to be 90% confident of detecting a difference of 5 percentage points or more in either direction at the 5% level of significance.

Solution

(*a*) Test coverage rate	40%
(*b*) Anticipated coverage rate	35% or 45%
(*c*) Level of significance	5%
(*d*) Power of the test	90%
(*e*) Alternative hypothesis (two-sided test)	coverage rate $\neq 40\%$

Table 3c (page 31) shows that for $P_0=0.40$ and $|P_a-P_0|=0.05$ a sample size of 1022 would be needed.

Two-sample situations

Estimating the difference between two population proportions with specified absolute precision

Required information and notation

(*a*) Anticipated population proportions	P_1 and P_2
(*b*) Confidence level	$100(1-\alpha)\%$
(*c*) Absolute precision required on either side of the true value of the difference between the proportions (in percentage points)	d
(*d*) Intermediate value	$V = P_1(1-P_1) + P_2(1-P_2)$

For any value of d, the sample size required will be largest when both P_1 and P_2 are equal to 50%; therefore if it is not possible to estimate either population proportion, the "safest" choice of 0.5 should be used in both cases.

The value of V may be obtained directly from Table 4a (page 33) from the column corresponding to P_2 (or its complement) and the row corresponding to P_1 (or its complement).

Tables 4b and 4c (pages 34–35) present minimum sample sizes for confidence levels of 95% and 90%, respectively.

Example 9

What size sample should be selected from each of two groups of people to estimate a risk difference to within 5 percentage points of the true difference with 95% confidence, when no reasonable estimate of P_1 and P_2 can be made?

Solution

(*a*) Anticipated population proportions ("safest" choice)	50%, 50%
(*b*) Confidence level	95%
(*c*) Absolute precision	5 percentage points
(*d*) Intermediate value	0.50

Table 4b (page 34) shows that for $d = 0.05$ and $V = 0.50$ a sample size of 769 would be needed in each group.

Example 10

In a pilot study of 50 agricultural workers in an irrigation project, it was observed that 40% had active schistosomiasis. A similar pilot study of 50 agricultural workers not employed on the irrigation project demonstrated that 32% had active schistosomiasis. If an epidemiologist would like to carry out a larger study to estimate the schistosomiasis risk difference to

within 5 percentage points of the true value with 95% confidence, how many people must be studied in each of the two groups?

Solution

(*a*) Anticipated population proportions	40%, 32%
(*b*) Confidence level	95%
(*c*) Absolute precision	5 percentage points
(*d*) Intermediate value	0.46

Table 4b (page 34) shows that for $d = 0.05$ and $V = 0.46$ a sample size of 707 would be needed in each group.

Hypothesis tests for two population proportions

This section applies to studies designed to test the hypothesis that two population proportions are equal. For studies concerned with very small proportions, see Example 13.

Required information and notation

(*a*) Test value of the difference between the population proportions under the null hypothesis	$P_1 - P_2 = 0$
(*b*) Anticipated values of the population proportions	P_1 and P_2
(*c*) Level of significance	$100\alpha\%$
(*d*) Power of the test	$100(1-\beta)\%$
(*e*) Alternative hypothesis: either	$P_1 - P_2 > 0$ or $P_1 - P_2 < 0$ (for one-sided test)
or	$P_1 - P_2 \neq 0$ (for two-sided test)

Tables 5a–h (pages 36–41) present minimum sample sizes for a level of significance of 5%, powers of 90% and 80%, both one-sided and two-sided tests, and the special case of very small proportions. Tables 5e–h should be used whenever the proportion under consideration is less than 5%.[1]

Example 11

It is believed that the proportion of patients who develop complications after undergoing one type of surgery is 5% while the proportion of patients who develop complications after a second type of surgery is 15%. How large should the sample size be in each of the two groups of patients if an investigator wishes to detect, with a power of 90%, whether the second procedure has a complication rate significantly higher than the first at the 5% level of significance?

Solution

(*a*) Test value of difference in complication rates	0%
(*b*) Anticipated complication rates	5%, 15%
(*c*) Level of significance	5%
(*d*) Power of the test	90%
(*e*) Alternative hypothesis (one-sided test)	risk difference $(P_1 - P_2) < 0\%$

[1] For further discussion of small proportions, see Lemeshow, S. et al., *Adequacy of sample size in health studies* (Chichester, John Wiley, 1990; published on behalf of the World Health Organization).

Table 5a (page 36) shows that for $P_1 = 0.05$ and $P_2 = 0.15$ a sample size of 153 would be needed in each group.

Example 12

In a pilot survey in a developing country, an epidemiologist compared a sample of 50 adults suffering from a certain neurological disease with a sample of 50 comparable control subjects who were free of the disease. Thirty of the subjects with the disease (60%) and 25 of the controls (50%) were involved in fishing-related occupations. If the proportion of people involved in fishing-related occupations in the entire population is similar to that observed in the pilot survey, how many subjects should be included in a larger study in each of the two groups if the epidemiologist wishes to be 90% confident of detecting a true difference between the groups at the 5% level of significance?

Solution

(*a*) Test value of difference between proportions involved in fishing-related occupations	0%
(*b*) Anticipated proportions involved in fishing-related occupations	60%, 50%
(*c*) Level of significance	5%
(*d*) Power of the test	90%
(*e*) Alternative hypothesis (two-sided test)	risk difference $\neq 0\%$

The required sample size is obtained from Table 5c (page 38) from the column corresponding to the smallest of P_1, P_2 and their complements and the row corresponding to $|P_2 - P_1|$. In this case, for $(1-P_1)=0.40$ and $|P_2 - P_1| = 0.10$, the required sample size would be 519 in each group.

Example 13

Two communities are to participate in a study to evaluate a new screening programme for early identification of a particular type of cancer. In one community the screening programme will include all adults over the age of 35, whereas in the second community the procedure will not be used at all. The annual incidence of the type of cancer under study is 50 per 100 000 ($=0.0005$) in an unscreened population. A drop in the rate to 20 per 100 000 ($=0.0002$) would justify using the procedure on a widespread basis. How many adults should be included in the study in each of the two communities if the investigators wish to have an 80% probability of detecting a drop in the incidence of this magnitude at the 5% level of significance?

Solution

(*a*) Test difference in cancer rates	0%
(*b*) Anticipated cancer rates	0.05%, 0.02%
(*c*) Level of significance	5%
(*d*) Power of the test	80%
(*e*) Alternative hypothesis (one-sided test)	risk difference $(P_1 - P_2) > 0\%$

Table 5f (page 40) shows that for $P_1 = 0.0005$ and $P_2 = 0.0002$ a sample size of 45 770 would be needed in each group.

Case–control studies

Examples 14 and 15 concern the odds ratio, which is the ratio of the odds of occurrence of an event in one set of circumstances to the odds of its occurrence in another. For example, if the "event" is a disease, people with and without the disease may be classified with respect to exposure to a given variable:

	Exposed	Unexposed
Disease	a	b
No disease	c	d

The odds ratio is then ad/bc.

Estimating an odds ratio with specified relative precision

Required information and notation

(*a*) Two of the following should be known:	
• Anticipated probability of "exposure" for people with the disease $[a/(a+b)]$	P_1^*
• Anticipated probability of "exposure" for people without the disease $[c/(c+d)]$	P_2^*
• Anticipated odds ratio	OR
(*b*) Confidence level	$100(1-\alpha)\%$
(*c*) Relative precision	ε

When the number of people in the population who are affected by the disease is small relative to the number of people unaffected:

$$c \approx (a+c)$$

and

$$d \approx (b+d).$$

In this case, therefore, the probability of "exposure" given "no disease" (P_2^*) is approximated by the overall exposure rate.

Tables 6a–h (pages 42–49) present minimum sample sizes for confidence levels of 95% and 90% and relative precisions of 10%, 20%, 25% and 50%.

For determining sample size from Table 6 when $OR \geqslant 1$, the values of both P_2^* and OR are needed. Either of these may be calculated, if necessary,

provided that P_1^* is known:

$$OR = [P_1^*/(1-P_1^*)]/[P_2^*/(1-P_2^*)]$$

and

$$P_2^* = P_1^*/[OR(1-P_1^*)+P_1^*].$$

If $OR < 1$, the values of P_1^* and $1/OR$ should be used instead.

Example 14

In a defined area where cholera is posing a serious public health problem, about 30% of the population are believed to be using water from contaminated sources. A case–control study of the association between cholera and exposure to contaminated water is to be undertaken in the area to estimate the odds ratio to within 25% of the true value, which is believed to be approximately 2, with 95% confidence. What sample sizes would be needed in the cholera and control groups?

Solution

(*a*) Anticipated probability of "exposure" given "disease"	?
Anticipated probability of "exposure" given "no disease" (approximated by overall exposure rate)	30%
Anticipated odds ratio	2
(*b*) Confidence level	95%
(*c*) Relative precision	25%

Table 6c (page 44) shows that for $OR = 2$ and $P_2^* = 0.3$ a sample size of 408 would be needed in each group.

Hypothesis tests for an odds ratio

This section outlines how to determine the minimum sample size for testing the hypothesis that the population odds ratio is equal to one.

Required information and notation

(*a*) Test value of the odds ratio under the null hypothesis	$OR_0 = 1$
(*b*) Two of the following should be known:	
• Anticipated probability of "exposure" for people with the disease $[a/(a+b)]$	P_1^*
• Anticipated probability of "exposure" for people without the disease $[c/(c+d)]$	P_2^*
• Anticipated odds ratio	OR_a
(*c*) Level of significance	$100\alpha\%$
(*d*) Power of the test	$100(1-\beta)\%$
(*e*) Alternative hypothesis (for two-sided test)	$OR_a \neq OR_0$

Tables 7a and 7b (pages 50–51) present minimum sample sizes for a level of significance of 5% and powers of 90% and 80% in two-sided tests.

For determining sample size from Table 7 when $OR_a > 1$, the values of both P_2^* and OR are needed. Either of these may be calculated, if necessary,

provided that P_1^* is known:

$$OR_a = [P_1^*/(1-P_1^*)]/[P_2^*/(1-P_2^*)]$$

and

$$P_2^* = P_1^*/[OR_a(1-P_1^*) + P_1^*].$$

If $OR_a < 1$, the values of P_1^* and $1/OR_a$ should be used instead.

Example 15

The efficacy of BCG vaccine in preventing childhood tuberculosis is in doubt and a study is designed to compare the vaccination coverage rates in a group of people with tuberculosis and a group of controls. Available information indicates that roughly 30% of the controls are not vaccinated. The investigators wish to have an 80% chance of detecting an odds ratio significantly different from 1 at the 5% level. If an odds ratio of 2 would be considered an important difference between the two groups, how large a sample should be included in each study group?

Solution

(*a*)	Test value of the odds ratio	1
(*b*)	Anticipated probability of "exposure" given "disease"	?
	Anticipated probability of "exposure" given "no disease"	30%
	Anticipated odds ratio	2
(*c*)	Level of significance	5%
(*d*)	Power of the test	80%
(*e*)	Alternative hypothesis	odds ratio $\neq 1$

Table 7b (page 51) shows that for $OR = 2$ and $P_2^* = 0.30$ a sample size of 130 would be needed in each group.

Cohort studies

Estimating a relative risk with specified relative precision

Required information and notation

(*a*) Two of the following should be known:	
• Anticipated probability of disease in people exposed to the factor of interest	P_1
• Anticipated probability of disease in people not exposed to the factor of interest	P_2
• Anticipated relative risk	RR
(*b*) Confidence level	$100(1-\alpha)\%$
(*c*) Relative precision	ε

Tables 8a–h (pages 52–59) present minimum sample sizes for confidence levels of 95% and 90%, and levels of precision of 10%, 20%, 25% and 50%.

For determining sample size from Table 8 when $RR \geqslant 1$, the values of both P_2 and RR are needed. Either of these may be calculated, if necessary, provided that P_1 is known:

$$RR = P_1/P_2$$

and

$$P_2 = P_1/RR.$$

If $RR < 1$, the values of P_1 and $1/RR$ should be used instead.

Example 16

An epidemiologist is planning a study to investigate the possibility that a certain lung disease is linked with exposure to a recently identified air pollutant. What sample size would be needed in each of two groups, exposed and not exposed, if the epidemiologist wishes to estimate the relative risk to within 50% of the true value (which is believed to be approximately 2) with 95% confidence? The disease is present in 20% of people who are not exposed to the air pollutant.

Solution

(*a*) Anticipated probability of disease given "exposure"	?
Anticipated probability of disease given "no exposure"	20%
Anticipated relative risk	2
(*b*) Confidence level	95%
(*c*) Relative precision	50%

Table 8d (page 55) shows that for $RR = 2$ and $P_2 = 0.20$ a sample size of 44 would be needed in each group.

Hypothesis tests for a relative risk

This section outlines how to determine the minimum sample size for testing the hypothesis that the population relative risk is equal to one.

Required information and notation

(*a*) Test value of the relative risk under the null hypothesis	$RR_0 = 1$
(*b*) Two of the following should be known:	
• Anticipated probability of disease in people exposed to the variable	P_1
• Anticipated probability of disease in people not exposed to the variable	P_2
• Anticipated relative risk	RR_a
(*c*) Level of significance	$100\alpha\%$
(*d*) Power of the test	$100(1-\beta)\%$
(*e*) Alternative hypothesis (for two-sided test)	$RR_a \neq RR_0$

Tables 9a–c (pages 60–62) present minimum sample sizes for a level of significance of 5% and powers of 90%, 80% and 50% in two-sided tests.

For determining sample size from Table 9 when $RR_a > 1$, the values of both P_2 and RR_a are needed. Either of these may be calculated, if necessary, provided that P_1 is known:

$$RR_a = P_1/P_2$$

and

$$P_2 = P_1/RR_a.$$

If $RR_a < 1$, the values of P_1 and $1/RR_a$ should be used instead.

Example 17

Two competing therapies for a particular cancer are to be evaluated by a cohort study in a multicentre clinical trial. Patients will be randomized to either treatment A or treatment B and will be followed for 5 years after treatment for recurrence of the disease. Treatment A is a new therapy that will be widely used if it can be demonstrated that it halves the risk of recurrence in the first 5 years after treatment (i.e. $RR_a = 0.5$); 35% recurrence is currently observed in patients who have received treatment B. How many patients should be studied in each of the two treatment groups if the investigators wish to be 90% confident of correctly rejecting the null hypothesis ($RR_0 = 1$), if it is false, and the test is to be performed at the 5% level of significance?

Solution

(*a*) Test value of the relative risk	1
(*b*) Anticipated probability of recurrence given treatment A	?
Anticipated probability of recurrence given treatment B	35%
Anticipated relative risk	0.5
(*c*) Level of significance	5%
(*d*) Power of the test	90%
(*e*) Alternative hypothesis	relative risk $\neq 1$

Table 9a (page 60) shows that for $RR_a = 0.5$ ($1/RR_a = 2$) and $P_2 = 0.35$ ($P_1 = 0.175$) a sample size of 135 would be needed in each group (figure obtained by interpolation; the exact sample size is 131 by computation).

Lot quality assurance sampling

Accepting a population prevalence as not exceeding a specified value

This section outlines how to determine the minimum sample size that should be selected from a given population so that, if a particular characteristic is found in no more than a specified number of sampled individuals, the prevalence of the characteristic in the population can be accepted as not exceeding a certain value.

Required information and notation

(*a*) Anticipated population prevalence	P
(*b*) Population size	N
(*c*) Maximum number of sampled individuals showing characteristic	d^*
(*d*) Confidence level	$100(1-\alpha)\%$

Tables 10a–j (pages 63–68) present minimum sample sizes for confidence levels of 95% and 90% and values of d^* of 0–4.

Example 18 In a school of 2500 children, how many children should be examined so that if no more than two are found to have malaria parasitaemia it can be concluded, with 95% confidence, that the malaria prevalence in the school is no more than 10%?

Solution

(*a*) Anticipated population prevalence	10%
(*b*) Population size	2500
(*c*) Maximum number of malaria cases in the sample	2
(*d*) Confidence level	95%

Table 10c (page 64) shows that for $P=0.10$ and $N=2500$ a sample size of 61 children would be needed.

Decision rule for "rejecting a lot"

This section applies to studies designed to test whether a "lot" (a sampled population) meets a specified standard. The null hypothesis is that the proportion of individuals in the population with a particular characteristic is equal to a given value, and a one-sided test is set up such that the lot is accepted as meeting the specified standard *only* if the null hypothesis can

be rejected. For this purpose a "threshold value" of individuals with the characteristic (d^*) is computed as a basis for a decision rule; if the number of sampled individuals found to possess the characteristic does not exceed the threshold, the null hypothesis is rejected (and the lot is accepted), whereas if the threshold is exceeded, the lot is rejected.

Required information and notation

(*a*) Test value of the population proportion under the null hypothesis	P_0
(*b*) Anticipated value of the population proportion	P_a
(*c*) Level of significance	$100\alpha\%$
(*d*) Power of the test	$100(1-\beta)\%$

Tables 11a–c (pages 69–71) present minimum sample sizes for a level of significance of 5% and powers of 90%, 80% and 50% in one-sided tests.

Example 19

In a large city, the local health authority aims at achieving a vaccination coverage of 90% of all eligible children. In response to concern about outbreaks of certain childhood diseases in particular parts of the city, a team of investigators from the health authority is planning a survey to identify areas where vaccination coverage is 50% or less so that appropriate action may be taken. How many children should be studied, as a minimum, in each area and what threshold value should be used if the study is to test the hypothesis that the proportion of children *not* vaccinated is 50% or more, at the 5% level of significance? The investigators wish to be 90% sure of recognizing areas where the target vaccination coverage has been achieved (i.e. where only 10% of children have not been fully vaccinated).

Solution

(*a*) Test value of the population proportion	50%
(*b*) Anticipated value of the population proportion	10%
(*c*) Level of significance	5%
(*d*) Power of the test	90%

Because the mistake of accepting groups of children as adequately vaccinated, when in fact the coverage is 50% or less, is the more important, $P_0 = 0.50$ and $P_a = 0.10$. Table 11a (page 69) shows that in this case a sample size of 10 and a threshold value of 2 should be used.

Therefore, a sample of 10 children should be taken from each of the areas under study. If more than 2 children in a sample are found not to have been adequately vaccinated, the lot (the sampled population) should be "rejected", and the health authority may take steps to improve vaccination coverage in that particular area. If, however, only 2 (or fewer) children are found to be inadequately vaccinated, the null hypothesis should be rejected and the group of children may be accepted as not being of immediate priority for an intensified vaccination campaign.

Incidence-rate studies

Estimating an incidence rate with specified relative precision

Required information and notation

(*a*) Relative precision	ε
(*b*) Confidence level	$100(1-\alpha)\%$

Table 12 (page 72) presents minimum sample sizes for confidence levels of 99%, 95% and 90%.

Example 20

How large a sample of patients should be followed up if an investigator wishes to estimate the incidence rate of a disease to within 10% of its true value with 95% confidence?

Solution

(*a*) Relative precision	10%
(*b*) Confidence level	95%

Table 12 shows that for $\varepsilon = 0.10$ and a confidence level of 95% a sample size of 385 would be needed.

Hypothesis tests for an incidence rate

This section applies to studies designed to test the hypothesis that the incidence rate of a characteristic is equal to a particular value.

Required information and notation

(*a*) Test value of the population incidence rate under the null hypothesis		λ_0
(*b*) Anticipated value of the population incidence rate		λ_a
(*c*) Level of significance		$100\alpha\%$
(*d*) Power of the test		$100(1-\beta)\%$
(*e*) Alternative hypothesis: either		$\lambda_a > \lambda_0$ or $\lambda_a < \lambda_0$ (for one-sided test)
	or	$\lambda_a \neq \lambda_0$ (for two-sided test)

Tables 13a–d (pages 73–76) present minimum sample sizes for a level of significance of 5%, powers of 90% and 80% and both one-sided and two-sided tests.

Example 21

On the basis of a 5-year follow-up study of a small number of people, the annual incidence rate of a particular disease is reported to be 40%. What minimum sample size would be needed to test the hypothesis that the population incidence rate is 40% at the 5% level of significance? It is desired that the test should have a power of 90% of detecting a true annual incidence rate of 50% and the investigators are interested in rejecting the null hypothesis only if the true rate is greater than 40%.

Solution

(*a*) Test value of the incidence rate	40%
(*b*) Anticipated incidence rate	50%
(*c*) Level of significance	5%
(*d*) Power of the test	90%
(*e*) Alternative hypothesis (one-sided test)	incidence rate > 40%

Table 13a (page 73) shows that for $\lambda_0 = 0.40$ and $\lambda_a = 0.50$ a minimum sample size of 169 would be needed.

Hypothesis tests for two incidence rates in follow-up (cohort) studies

This section applies to studies designed to test the hypothesis that the true incidence rates of a disorder or characteristic in two groups of individuals are equal. Subjects either have a common date of entry into the study and are followed up until they develop the characteristic in question or cannot be followed up any more (Example 22), or are inducted into the study as they become available but are followed up only until a specified date (Example 23).

Required information and notation

(*a*) Test value of the difference between the population incidence rates under the null hypothesis	$\lambda_1 - \lambda_2 = 0$
(*b*) Anticipated values of the incidence rates	λ_1 and λ_2
(*c*) Level of significance	$100\alpha\%$
(*d*) Power of the test	$100(1-\beta)\%$
(*e*) Alternative hypothesis: either	$\lambda_1 - \lambda_2 > 0$ or $\lambda_1 - \lambda_2 < 0$ (for one-sided test)
or	$\lambda_1 - \lambda_2 \neq 0$ (for two-sided test)
(*f*) Duration of study (if fixed)	T

If the study is terminated at a fixed point in time, before all the subjects have necessarily experienced the end-point of interest, the observations are said to be *censored*. The values of λ then have to be modified according to the formula

$$f(\lambda) = \lambda^3 T/(\lambda T - 1 + e^{-\lambda T})$$

as in Example 23.

Tables 14a–d (pages 77–80) present minimum sample sizes for a level of significance of 5%, powers of 90% and 80% and both one-sided and two-

sided tests, when *the duration of the study is not fixed* and the two groups studied are of equal size. No tables are given for studies of fixed duration because too many parameters are involved to permit easy tabulation.

Example 22

As part of a study of the long-term effect of noise on workers in a particularly noisy industry, it is planned to follow up a cohort of people who were recruited into the industry during a given period of time and to compare them with a similar cohort of individuals working in a much quieter industry. Subjects will be followed up for the rest of their lives or until their hearing is impaired. The results of a previous small-scale survey suggest that the annual incidence rate of hearing impairment in the noisy industry may be as much as 25%. How many people should be followed up in each of the groups (which are to be of equal size) to test the hypothesis that the incidence rates for hearing impairment in the two groups are the same, at the 5% level of significance and with a power of 80%? The alternative hypothesis is that the annual incidence rate for hearing impairment in the quieter industry is not more than the national average of about 10% (for people in the same age range), whereas in the noisy industry it differs from this.

Solution

(*a*) Test value of the difference in incidence rates	0
(*b*) Anticipated incidence rates	25% and 10%
(*c*) Level of significance	5%
(*d*) Power of the test	80%
(*e*) Alternative hypothesis (two-sided test)	$\lambda_1 \neq \lambda_2$
(*f*) Duration of study	not applicable

Table 14d (page 80) shows that for $\lambda_1 = 0.25$ and $\lambda_2 = 0.10$ a sample size of 23 would be required in each group.

Example 23

A study similar to that outlined in Example 22 is to be undertaken, but the duration of the study will be limited to 5 years. How many subjects should be followed up in each group?

Solution

(*a*) Test value of the difference in incidence rates	0
(*b*) Anticipated incidence rates	25% and 10%
(*c*) Level of significance	5%
(*d*) Power of the test	80%
(*e*) Alternative hypothesis (two-sided test)	$\lambda_1 \neq \lambda_2$
(*f*) Duration of study	5 years

The values of λ must be modified according to the formula for $f(\lambda)$ given on page 18:

$$f(\bar{\lambda} = 0.175) = 0.0918 \text{ where } \bar{\lambda} = (\lambda_1 + \lambda_2)/2$$
$$f(\lambda_1 = 0.25) = 0.1456$$
$$f(\lambda_2 = 0.10) = 0.0469.$$

The appropriate sample size formula is

$$n_1 = \{z_{1-\alpha/2}\sqrt{[(1+k)f(\bar{\lambda})]} + z_{1-\beta}\sqrt{[kf(\lambda_1)+f(\lambda_2)]}\}^2/(\lambda_1 - \lambda_2)^2,$$

where k is the ratio of the sample size for the second group of subjects (n_2) to that for the first group (n_1) (in this example $k = 1$).

Thus

$$n_1 = \{1.96\sqrt{[2(0.0918)]} + 0.842\sqrt{(0.1456 + 0.0469)}\}^2/(0.25 - 0.10)^2$$
$$= 1.462/0.023$$
$$= 65.0.$$

A sample size of 65 would therefore be needed for each group.

For a one-sided test the corresponding sample size formula is

$$n_1 = \{z_{1-\alpha}\sqrt{[(1+k)f(\bar{\lambda})]} + z_{1-\beta}\sqrt{[kf(\lambda_1) + f(\lambda_2)]}\}^2/(\lambda_1 - \lambda_2)^2.$$

Definitions of commonly used terms

The brief definitions listed here are intended to serve only as reminders for the reader. Fuller explanations of statistical terms and a discussion of the statistical theory relevant to sample size determination are to be found in Lemeshow, S. et al., *Adequacy of sample size in health studies* (Chichester, John Wiley, 1990; published on behalf of the World Health Organization).

α	The significance level of a test: the probability of rejecting the null hypothesis when it is true (or the probability of making a Type I error).
β	The probability of failing to reject the null hypothesis when it is false (or the probability of making a Type II error).
Case–control studies	Studies in which subjects are selected on the basis of their status with respect to a given characteristic (such as the presence of a disease); the "cases" show the characteristic and the "controls" do not. Both groups are studied with respect to their prior and current exposure to suspected risk factors.
Cluster sampling	A sampling process in which sampling units are made up of clusters or groups of *study units*.
Cohort studies	Studies in which subjects are selected with respect to the presence and absence of a characteristic (such as exposure to a given factor) suspected of being associated with the particular outcome of interest (for example a disease). Both groups of subjects are followed up for development of the outcome.
Confidence level	The probability that an estimate of a population parameter is within certain specified limits of the true value; commonly denoted by "$1-\alpha$".
Design effect	In *cluster sampling*, the design effect is an indication of the variation due to clustering. It is estimated by the ratio of the variance when *cluster sampling* is used to the variance when *simple random sampling* is used.
Incidence rate	The number of specific events (for example new cases of a disease) occurring in a specified population per unit time.
Lot quality assurance sampling	Sampling techniques, with industrial origins, designed to ascertain whether batches of items meet specified standards.
Null hypothesis	A statement concerning the value of a population parameter. It is the hypothesis under test in a test of significance, for example the hypothesis that an observed difference is entirely due to sampling error.

Odds ratio — The ratio of the odds of occurrence of an event in one set of circumstances to the odds of its occurrence in another (see also page 9).

One-sided test — In hypothesis testing, when the difference being tested is directionally specified beforehand (for example when $X_1 < X_2$, but not $X_1 > X_2$, is being tested against the null hypothesis $X_1 = X_2$).

Population proportion — The proportion of individuals in a population possessing a given characteristic.

Power of a test — The probability of correctly rejecting the null hypothesis when it is false; commonly denoted by "$1-\beta$".

Precision — A measure of how close an estimate is to the true value of a population parameter. It may be expressed in absolute terms or relative to the estimate.

Prevalence — The number of cases of a disease (or people with a particular characteristic) existing in a specified population at a given point in time.

Relative risk — The ratio of the risk (probability) of an outcome (for example disease or death) among people exposed to a given factor to the risk among people not exposed.

Significance level — See definition of α.

Simple random sampling — Sampling procedure in which every *study unit* has the same chance of being selected and every sample of the same size has the same chance of being chosen.

Study units — The individual members of a population whose characteristics are to be measured.

Two-sided test — In hypothesis testing, when the difference being tested for significance is not directionally specified beforehand (for example when the test takes no account of whether $X_1 > X_2$ or $X_1 < X_2$).

$z_{1-\alpha}$, $z_{1-\alpha/2}$ and $z_{1-\beta}$ — Represent the number of standard errors from the mean; $z_{1-\alpha}$ and $z_{1-\alpha/2}$ are functions of the confidence level and $z_{1-\beta}$ is a function of the power of the test.