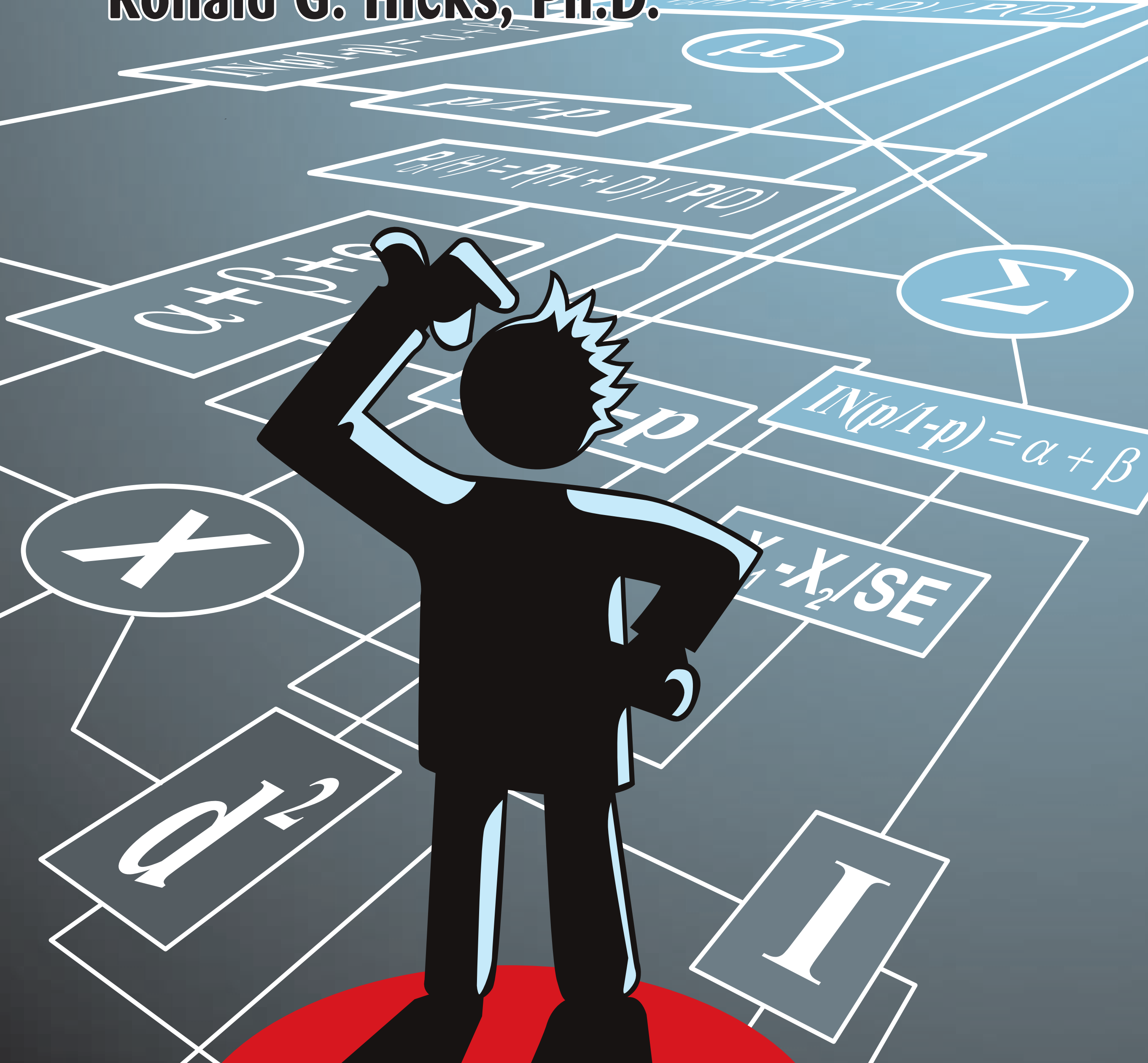
Primer for Health Research:
Design & Basic Statistics

Ronald G. Hicks, Ph.D.



ISBN 978-0-9808000-0-5 PBk

This book has been produced for the Publisher by Bookbound

To order your copy, visit
www.primerforhealthresearch.com.au

Primer for Health Research: Design & Basic Statistics

Ronald G. Hicks, Ph.D.

Spinoza
Publishers

Dedication

This book is dedicated to Wanda for her encouragement and support in creating the time to write; to Annika, Stefan and their progeny; and in memory of Amour for his unconditional love and his example of remaining physically robust, even when the odds were stacked against him.

Preface

Statistics is a tool used to formalise and standardise procedures for testing hypotheses – this involves statistical inference. Statistical inference draws conclusions encompassing a large number of events thereby introducing order from the evidence contained within the data. However differences can and do occur by chance. How is chance separated from some systematic order? Chance exists within a range of possibilities, once outside that range of fluctuation, the detection of systematic order becomes feasible.

Biostatistics is a combination of physiology and statistics. Knowledge of both is an important ingredient in the planning and conducting of health research. After attending numerous professional conferences on health research, the overall conclusion drawn from the presentations was that of barely adequate experimental design. Considerable forethought could have yielded better results, and more elaborate conclusions. A well-designed experiment developed at the very start is the key to success. No amount of remediation will make the experiment and/or data any better. Think first. Then collect data. Considerable emphasis is placed upon description and percentages with little or no understanding of foundational concepts. This book establishes the basic requirements for an adequate bioresearch design and once that data has been collected, how to separate chance from systematic resultants via statistical tests.

Follies in experimental design, kinds of data collected and data not collected, fly by the seat of your trousers attitude, can totally destroy any project, large or small, and with the loss of considerable human time and effort. Moreover, people working on the project will be disheartened from attempting other projects. Success breeds success. Trial and error is a hard way to learn; whereas, developing a framework of reference will increase the probability of success. All too often while attending professional conferences, the speaker will announce a percentage difference between 'A' and 'B'. Quick scanning the power point data reveals that even the percentage differences have been miscalculated. Percentage difference, even when calculated correctly, reveals little or nothing. Is the difference chance or effect?

Another folly is the nature of the data itself, nominal, ordinal, and/or ratios scales have been mixed into the same data column. For most statisticians ninety percent of the job is cleaning up the data including throwing away bales of unusable data. Success is in the details. This book will provide a small but important introduction to biostatistics and thereby optimistically increase the quality of conference presentations and professional papers, not to mention published results.

The author would like to thank Dr Ray Haning and Ruth Burrell for their considerable contributions.

Contents

Randomisation

Biostatistics is a combination of physiology and statistics. Knowledge of both is an important ingredient in the planning and conducting of health research. After attending numerous professional conferences on health research, the overall conclusion from the presentations was that of barely adequate experimental design. Considerable forethought could have yielded better results, and more elaborate conclusions. A well-designed experiment developed at the very start is the key to success. No amount of remediation will make the experiment and/or data any better. Considerable emphasis is placed upon description and percentages with little or no understanding of a foundation concept of randomness. A simple discussion of the randomness, extremely oversimplified, follows.

Randomness

Randomness, at the basic core of this concept, states, that all events have an equal chance of occurring. If a very large bowl contains ten thousand coloured balls of all hues and saturations, and you only wanted the most saturated black ball. What is your random chance of picking out the black ball from the ten thousand coloured balls? Simply expressed, the chance to draw the black ball is one chance in ten thousand (or rewritten in mathematical terms is 1/10000, simplified to 0.0001, or 1.0 (-10^4). Now the question becomes what are the chances of drawing out a red ball? Again, the answer is the same, 0.0001. All the balls have an equal opportunity of selection from the entire collection of balls. Now comes a point of needless confusion. If you have drawn the desired black ball, but now want to draw-out the most saturated red ball, do you replace the black ball back into bowl with all the other balls? Yes. If you do not, the chances have changed from 1/10,000 to 1/9,999. In an experiment, all subjects must have an equal chance of being drawn, i.e., the chance of selection for all subjects must be the same. If not, then the selection process is biased. The chance of selection has been altered with each draw. Keeping a constant number in the pool for selection is an important undertaking in the randomising process.

Note: *You are in a biology class of thirty students, and some of the students are being chosen to live for three months in a country of their selection for physiological studies. Jane chose to go to Italy. John selected to go to Sweden, and Mark is going to Germany. As more students are chosen, does your chance of studying in another country increase? Absolutely, yes, from 1/30 originally, with Jane's selection 1/29 remains, with John's selection 1/28 remain, with Mark's selection 1/27 remains. Your chances of being selected are improving with each new selection. Is this truly a random selection process? NO! The size of the original pool of thirty students is constantly being changed; the pool of students is being reduced with each new selection.*

Description

Description of an event or process in the medical/health is common, often reminiscent of six blind people examining a platypus. One person is provided the head, another person the tail, and the four remaining people are provided with the legs. The platypus is female, minus the poisonous back-leg spurs. One person describes the animal as being flat made of flexible rubbery tissue, another person partially agrees with the description. The other four people describe the animal as a webbed, soft, malleable tissue. Naturally, all are correct, and all are limited in their perspective of the nature of a platypus, depending upon what part of the animal is being examined.

Description can be helpful but is limited in conveying overall information. Commonly, varying forms of description are used in parlance, sometimes called 'naturalistic observations', sometimes 'qualitative analyses'. Description can be a starting process in the scientific method, but only, and at best, a starting place.

Percentage

Poor percentages are grossly over emphasised as a means of understanding some kinds of data. Percentages alone do not convey much, if any, meaningful information. Is the percentage, or change in percentage, outside random fluctuations found in a selected environment? This introduces an important topic, when random fluctuations are not random; rather the fluctuations are a function of some treatment, some effect, or some manipulation.

Level of Confidence

The underlying question is random fluctuates responsible for the percentage differences? Random fluctuations gain importance when the change, the odds of an event, treatment, or manipulation exceeds an established level of confidence. Let us pretend we are at a horse race track with a large group of acquaintances. You would like to place a modest bet for the winner of any of the races and start to listen to your acquaintances as they discuss the chances of a particular horse winning the next race. You ask how likely is the chosen horse is to win. You learn that, the horse has a 60% chance of winning, or this horse would win a race 60 times in 100 races. The odds of winning are not good enough for you, so you sit down and enjoy the race. Your acquaintances start discussing the next race. Someone in the group relates they have inside information that horse Plug is going to win. Again you ask how likely the chosen horse, Plug, is to win. The replay this time is 85% chance of winning, or winning 85 times out of 100 races. Not good enough, so you sit down and enjoy the race. Again your acquaintances discuss the next race and one in the group relates that the horse Phar Lap, is going to win, hands down, no problems. Being consistent you again ask, what are Phar Lap's chances of winning the next race? The answer this time is 95% certain that this horse is going to win, or win 95 races out of 100 attempts. This time you bet. You have a 95% chance Phar Lap will win, a 0.95 level of confidence of this win, but still mindful on this particular race, on this particular day, this race could be one to the five times in a hundred that aspiring Phar Lap will lose. Now, you are willing to take the risk this is not one of those five times that Phar Lap will lose.

In scientific notation the other side of the chance ratio is used. You, who bet your money, have a 5% chance of losing that money, or a probability 0.05 that the predicted outcome will occur by some random fluctuation and not some treatment, or effect, or manipulation, or in this case a whisper. You will bet your money with a 0.05 level of significance the outcome that your horse will win. Now back to percentages.

Note: *Throughout the state health system, two hundred clients are using cardiac beta-blocker A, and another two hundred clients are using beta-blocker B. After a year all the clients have their heart rate tested. The clients receiving beta-blocker A have an average (Mean) rate of 87, while the clients using beta-blocker B have an average (Mean) rate of 78. One question, from many possible questions, did using beta-blocker B yield better results than beta-blocker A? A difference of nine percent exists between the usages of the different beta-blockers. A statistical test reveals a level of significance of 0.025. Is there an acceptable scientific difference to determine performance difference between the two beta-blockers? Yes! Only about 2.5 times in a hundred will the difference, 87 versus 78 will occur by chance. That level of chance is scientifically acceptable. The different beta-blockers yielded significantly different performances in the reduction of heart rate.*

Percentages revisited

If you reported at a conference, people with a BMI (Body Mass Index) over 30 had a 56% increase in heart attacks compared with people with a BMI of under 28, fine. Now if you added that the chance of having a heart attack given the above results is significant at the 0.05 level of confidence, and these results could only occur 5 times in 100 by chance, then the scientific community is accepting the 56% increase of added risk. Now you have every one's attention. Your results could have occurred by random fluctuation five times in a hundred, but we are accepting your results are not one of those times, rather, the results are attributed to the differences in the BMI. The next question is, 'how do I determine if my percentage difference is significant?' This question can be answered by the use of a statistical test.

Conclusion

How many times have 'steering committees' made a complete and expensive hash out of collecting useless data, when some forethought could have produced tangible conclusions. Planning and thinking ahead has no substitute. Once data collection is started, an attempted retro-fix is too late. The remaining chapters will outline how to avoid this problem; as well as, discuss in modest detail the use of statistical methods, namely: (1) how to make percentages, descriptions, differences and relationships more precise, (2) how to make predictions, (3) how to design a study, and (4) how to determine the degrees of confidence in the study's conclusions.

Readings

Randomness, Deborah J. Bennett.Harvard University Press, 1998 ISBN 0-674-10745-4.

Random Measures, 4th ed. Olav Kallenberg. Academic Press, New York, London; Akademie-Verlag, Berlin (1986). MR0854102

The Art of Computer Programming. Vol. 2: Seminumerical Algorithms, 3rd ed. by Donald E. Knuth, Reading, MA: Addison-Wesley, 1997. ISBN 0-201-89684-2

Fooled by Randomness, 2nd ed. by Nassim Nicholas Taleb. Thomson Texere, 2004. ISBN 1-58799-190-X

Exploring Randomness. Gregory Chaitin. Springer-Verlag London, 2001. ISBN 1-85233-417-7

Random. Kenneth Chan, includes a "Random Scale" for grading the level of randomness

The Question

How the question is phrased, often dictates the answer. Asking a socially acceptable question will often yield a socially acceptable answer, e.g., do you love your children? Do you beat your spouse, or children, or the dog? Questions like this will produce socially acceptable answers. Be clever and do not ask direct questions that will generally lead to socially acceptable answers.

Try not to ask binary questions, a favourite within the medical/health disciplines. Do you go to bed in-order to sleep – yes or no? Instead try asking; of how many times to you nod-off before retiring to bed? The binary question provides only two possible answers, yes or no, or in Boolean algebra a zero '0' or a '1'. Very few statistical tests exist to analysis binary data. Generally, dichotomous (like binary, exists only in two states) questions yielding binary data, providing only percentages, results and nothing more. Sometimes dichotomous questions are un-avoidable, like male or female, but in the main try to avoid this type of question.

Likert

A Likert scale is a considerably better alternative to a dichotomous outcome. The Likert scale asks for the answers to be graded along a continuum, e.g., one to about ten. How hard did you hit your spouse, one equals very lightly, graduated to ten for 'as hard a possible'. The numeric distance between two and three on the scale will probably not be the same distance between eight and nine on the scale. Each step on this Likert scale is discrete, but each step along the discrete scale is in order of severity, nine being more severe then eight, but the absolute degree of increased severity is unknown, therefore discrete categories of severity exist. Numerous statistical tests exist for data in discrete categories.

Note: *A typical medical question would be, "are you older than 55?" This is a dichotomous or binary question, the only resultant is either yes or no (zero or one). An improvement would be to categorise the answers, e.g. 0 to 9 years of age = category 1; 10 to 19 years of age = category 2; 20 to 29 years of age = category 3, etc. This would be categorical scale, e.g., similar to the Likert scale. If we ask for the person's date of birth, the data would be on continuous scale. Each change of the kinds of data collected, dictates further options. What numerical level is a BMI over 27?*

Kinds of Measurement

Measurement implies the assignment of numbers to observations so that the numbers are amenable to analysis in accordance with established rules.

1. **Nominal or Classification Data.** Already, we have come to understand all numbers are not created equal. Zero and one appear be familiar numbers, but the notation could just as easily be referenced "on" or "off". At this numeric level the notating symbols are referencing dissimilar categories, and are referred to as **Nominal data or Classification data**. One category could be oranges, the next sky, the next purple, and the next chimpanzees. All categories are uniquely different, mutually exclusive of each other. Only a few requirements for the Nominative categories of data, namely;
   a. Reflexive, where x = x for all values of x,
   b. Symmetrical, where if x = y, then y = x, and
   c. Transitive, where if x = y and y = z, then x = z. For this category of data, statistical tests can be accomplished using nonparametrics (no underlying assumption as to distribution of data); some being,
      i. the Chi-Square for testing hypothesis, statistical inference, and
      ii. Contingency Coefficient, "C", for determining the degree of association, relationship, correlation.
2. **Ordinal or Rating Scales.** The Likert scale would be an example of the Ordinal scale. Here the categories in the scale are related, of a similar kind. Some kind of relationship exists between the categories, like higher, more liked, more enjoyable. The mathematical symbol for this kind of relationship is (>), 'greater than' or the converse (<), 'less than'. Note, what does NOT pertain to the Ordinal scales is the absolute known distance between adjacent categories, how much greater than or how lesser than. Only a higher or lower relationship exists between categories. Ordinal scales have a few formal requirements; namely,
   a. Irreflexive, where any x is not necessarily x > x,
   b. Asymmetrical, where if x > y, then y >/ x,
   c. Transitive, where if x > y, and y > z, then x > z.

   Statistical tests do exist for Ordinal data, for example the use of the median for central tendency. Some statistics tests for Ordinal data use the median, for central tendency. The MEDIAN defines as a point on a scale of data where exactly half the data are above that point and the other half of the data are below that point. The median point is one of three measures of central tendency. The other two are the MEAN, the mathematical average, and the MODE is that point on a scale where the most scores reside. Hypothesis testing, determining if differences are random fluctuations or not, again nonparametric statistics, sometimes referred to as ranking or ordered statistics, including, but not inclusive, Mann-Whitney U test, Friedman two-way analysis of variance are the tests to be used. For determining relationships, degrees of association, correlation, include Spearman (r_S) and/or Kendall (r). The only assumption about Ordinal data is that the measured scores are drawn from an underlying continuous distribution, the distance from one point to an adjacent points is not known, only the direction of change.

3. **Interval scale.** Interval scale should be familiar to just about everyone. The Interval scale is similar to the Ordinal scale with one additional qualifier; the distance between adjacent numbers is constant. This distance is of a known size throughout the scale. Most health scientists try to use scales that are interval; the choice of statistical tests are more extensive and sometimes more powerful. Power efficiency and type-one and type-two errors will be discussed a little later.

   The Interval scale has all the properties of the nominal scale and the ordinal scale plus the distance between adjacent numbers are similar, sometimes referred to as being isomorphic. If the Interval scale has a true zero starting point, this sequence of numbers is a Ratio scale. Both the Interval and Ratio scales can use both nonparametric (does not assume a bell shaped curve as to data distribution) and parametric statistical tests (does assume a bell shaped curve of data distribution), e.g., “t-test”, “F-test”, and other parametric measures like means, standard deviations, and Pearson Product Correlations, for determining degrees of association or relationship. One exception exists, the Interval scale cannot use the coefficient of variation. In this case, a true zero must be known, the Ratio scale.

   Already some of the differences between parametric and nonparametric statistics have been established. Nonparametric statistics can be used to determine randomness for all scales; nominal, ordinal, interval and ratio; whereas, parametric statistics can only be used on data that is interval or ratio. Other differences between these two statistical families do exist.

Note: *The differences between the Mean, Median, and Mode, the 3Ms of the measurement of central tendency are calculated in the table below.*

Table 2.1: The 3Ms calculated

Distribution of Measures	Frequency of Entries	Mean	Median	Mode
98	1	98		
97	1	97		
95	2	190		
94	1	94		
93	3	279		
91	5	405	3/5=0.6*2=1.2	91
89	4	356	7	
88	1	88	3	
87	1	87	2	
86	1	86	1	
Sums =	20	1780		
		1780/20		
		X = 89	Md = 90.2	Mode=91

Calculating an interpolating median from grouped data:

1. Number of entries divided by 2 (N/2).
2. From the bottom count up until the interval containing the Median is found.
3. How many entries are needed from this interval to make N/2.
4. Divide the number needed by the number of entries within the interval.
5. Multiply answer in step 4 by the size of the interval.
6. Add the number obtained in step 5 to the lower limit of interval containing the Median.
7. Confirm the results by redoing the above steps but starting from the top.

Readings

Michell, J. (1986): Measurement scales and statistics: a clash of paradigms. *Psychological Bulletin*, *3*, 398-407.

Michell, J. (1997): Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, *88*, 355-383.

Michell, J. (1999): *Measurement in Psychology – A critical history of a methodological concept*. Cambridge: Cambridge University Press.

Michell, J. (2008): Is psychometrics pathological science? *Measurement – Interdisciplinary Research & Perspectives*, *6*, 7-24.

Narens, L. (1981a): A general theory of ratio scalability with remarks about the measurement-theoretic concept of meaningfulness. *Theory and Decision*, *13*, 1-70.

Stevens, S.S. (1946): On the theory of scales of measurement. *Science*, *103*, 677-680.

Stevens, S.S. (1951): Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49). New York: Wiley.

Stevens, S.S. (1975): *Psychophysics*. New York: Wiley.

ISBN 978-0-9808000-0-5
Paperback, 234x180mm
132pp; illustrated; glossary; index

Ordering within Australia & overseas

The cost of a single copy of the book is $55. Postage and handling cost for Australia is $5 per book. Postage and handling cost to anywhere else in the world is $20 per book. All prices are in Australian dollars and include GST where applicable.

Bulk and trade discounts available.

Payment can be made with a credit/debit card, via PayPal or with a cheque/money order (Australia only).

Credit/Debit card purchase

You can ring through your order on 1800 628 058 (Int'l +612 6568 4296) or fax it through to 02 6568 3960 (Int'l +612 6568 3960).

PayPal purchase

You can make a safe and secure payment straight from your bank account, credit/debit card.

Cheque or money order purchase

Australian customers only. To purchase with a cheque or money order, send your payment and postal details to
Bookbound Publishing, PO Box 233, Macksville, NSW 2447.

www.primerforhealthresearch.com.au

At last, we have an Australian introductory statistical text written specifically for health professionals working at the grass roots level.

Dr Ron Hicks has worked in the health statistics field, both in Australia and overseas for many years and has seen first hand the traps and pitfalls of well meaning health professionals trying to make sense of the avalanche of data collected in health today.

This text takes us from the rudimentary concepts of mean, median and mode to the application of numerous parametric and nonparametric tests and explains which ones to use and when to use them. The many examples given deal with common health related problems such as blood pressure, diabetes and wound healing, and we are able to work through the dilemma of whether to accept or reject data to ensure validity of the study. We then move on to evaluation, then how to predict patient outcomes using the data collected and finally how to manage multiple variables.

This statistical primer covers these and many other issues in a practical and straightforward manner and will be a valuable tool for health clinicians who are committed to improving outcomes for their patients and better forward planning decisions for their services.

Ruth Burrell RN
Former Quality, Accreditation and Program Manager

Spinoza Publishers

ISBN 978-0-9808000-0-5

9 780980 800005