

PHOTO BY HAFM JANSEN



Analyzing Quantitative Data

Topics covered in this chapter:

Basic analysis of survey data on violence against women Looking at associations between violence and other variables Assessing the validity of survey results Interpreting the results

nce the fieldwork has been completed, the next task, and arguably the most rewarding part of the research process, is to unravel the mysteries hidden within the data. This involves transforming endless pages of words, numbers, and codes into meaningful results to inform theory building or action.

One of the common limitations of survey research on violence is the tendency to present results mostly in the form of descriptive statistics, for example, percentages of women who are abused and characteristics of abuse. While these types of figures are certainly necessary, there is a broad range of statistical techniques for data analysis that allows researchers to explore the relationships between violence and other variables with greater depth and precision. In this chapter, we show how a variety of statistical techniques may be combined to enrich data analysis. Providing detailed guidance for advanced statistical analysis is beyond the scope of this manual, and we urge researchers needing further assistance to consult an experienced statistician during this stage.

BOX 12.1 STEPS FOR ANALYZING AND INTERPRETING THE DATA

- Inspect and clean the data matrix.
- Display frequencies of the variables.
- Make cross tabulations to check for inconsistencies.
- Check graphic distributions of relevant variables.
- Review basic hypothesis.
- Plan initial tables.
- Write down expected results.
- Carry out additional analysis.
- Illustrate results in tables and graphs.
- Interpret results and assess critically.
- Synthesize results in writing.

(Adapted from Persson and Wall, 2003.1)

BASIC ANALYSIS OF SURVEY DATA ON VIOLENCE **AGAINST WOMEN**

Look at the data matrix

Before initiating data analysis, it is important to make sure that the data are entered properly and that there are no obvious mistakes. There are several ways

TABLE 12.1 EXAMPLE OF A DATA MATRIX							
Record	Code MY132	Abused	Age 15	Numchild	Educ 4	Poverty	Urban
2	KJ423	1	29	3	2	0	2
3	MC341	2	31	2	2	1	2
4	KJ153	2	46	5	3	2	1
5	MI253	1	24	3	1	1	1
6	KU124	2	19	2	3	0	2

to do this. One way to check data entry is to look over the data matrix frame by frame and make sure that nothing is missing.

Another method is to make frequencies of the basic variables. Look for variables that seem to have been miscoded, and check for missing values on key variables such as age and education.

Finally, do cross-tabulations to look for obvious inconsistencies. For example, make sure that all the women who have answered questions about the characteristics of violence are also coded as having been abused, and that none of the women coded as not having children have answered the questions about children's age and schooling.

Whenever you find inconsistencies or missing data, you will probably need to consult the original questionnaire to find the correct answer. Be sure to keep a log of any corrections made to the data file. As soon as you have a fully updated and cleaned version of the data set, make several copies and keep them in a safe place!

This is certainly the most tedious part of data analysis. Doing it carefully can, however, save you time and anguish later on. Table 12.1 presents an example of what a data matrix looks like. Each row represents a different informant, and the columns represent different variables.

Prepare for data analysis

In preparing for the data analysis, you should return to your hypothesis and

review the relationships you are expecting to test. These can be presented in the form of empty tables that serve as your initial analysis plan. First, look at the frequencies of variables such as age and education to see how they are distributed. One way to look at distribution graphically is by means of a **histogram** (Figure 12.1), which measures the frequencies of observations in each category. This will be helpful later on in finding meaningful ways to recode the data into smaller categories, such as age group and educational attainment.

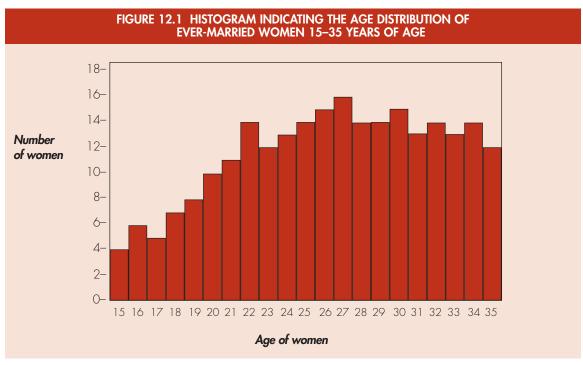
Different ways to describe the occurrence of violence

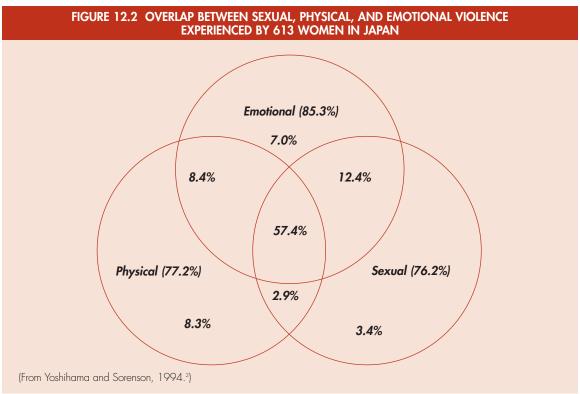
One of the most important findings of a prevalence survey on violence is the percentage of women who have been abused. However, there are many different ways to quantify and describe the frequency of violence, and you will need to analyze your data from several different angles to find the most meaningful way to describe your results. Following are some examples of the different ways that data on violence may be presented.

As we mentioned in Chapter 6, the prevalence of violence can be expressed as the percentage of women interviewed who have experienced violence during a specific period of time (for example, the last 12 months), or as the percentage of women who have ever experienced violence in their lives.

It is also useful to break down these figures according to the type of violence

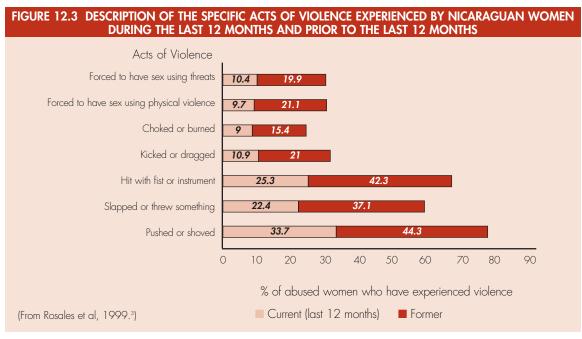


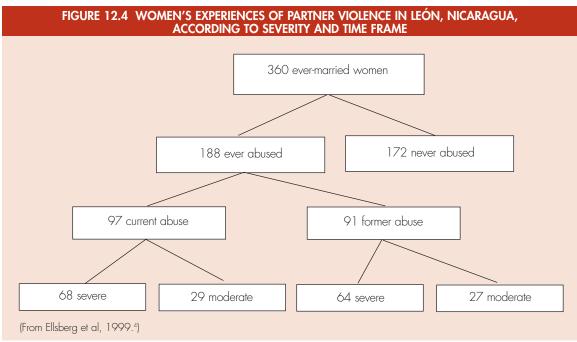




(physical, sexual, emotional, or economic) and the perpetrator (in other words, abuse by partner should be presented separately from abuse by other individuals).

It is further useful to look at the overlaps between different types of violence, for example, how many women have experienced both partner and nonpartner abuse, or how many women have experienced both sexual and physical violence. These can be illustrated using Venn diagrams. The diagram in Figure 12.2 was





produced by Yoshihama and Sorenson, and presents the types of abuse that were reported by women in Japan.² The percentages within parentheses represent the proportion of abused women who experience each type of abuse, whereas the numbers without parentheses represent the proportion of abused women who experienced specific combinations of violence. In this example, 57 percent of abused women

experienced all three types of abuse.

Most women who suffer abuse experience several different acts of violence, as well as repeated incidents of violence. To enable readers to visualize the range of abusive behaviors, it may be helpful to describe the percentage of abused women who have experienced specific acts of violence. These may be presented as lifetime experiences, or as in Figure 12.3, from the



1998 Nicaraguan Demographic and Health Survey, broken down by recent and former acts of violence experienced.

Figure 12.4 and Table 12.2 present other examples of how to present descriptive information on violence. Figure 12.4 shows that about half of the women reporting violence were abused during the last 12 months, and that 70 percent of both current and former violence was classified as severe. Table 12.2 shows that 60 percent of the women experiencing violence during the last year suffered more than one incident, and that 22 percent of women were severely abused more than five times in the last year. After completing this descriptive analysis, we recommend developing a scale that enables you to group women according to their experience of violence, for example: never abused, former moderate abuse, former severe abuse, current moderate abuse, and current severe abuse. This kind of detailed information about what kind of violence women suffer will

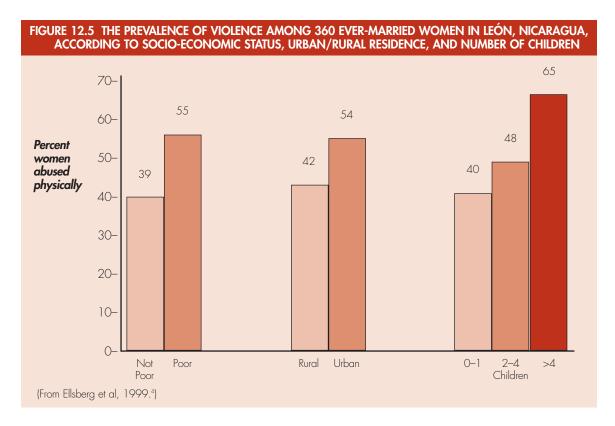
TABLE 12.2 FREQUENCY OF VIOLENCE IN LEÓN, NICARAGUA,
DURING THE LAST 12 MONTHS ACCORDING TO SEVERITY (N=97)

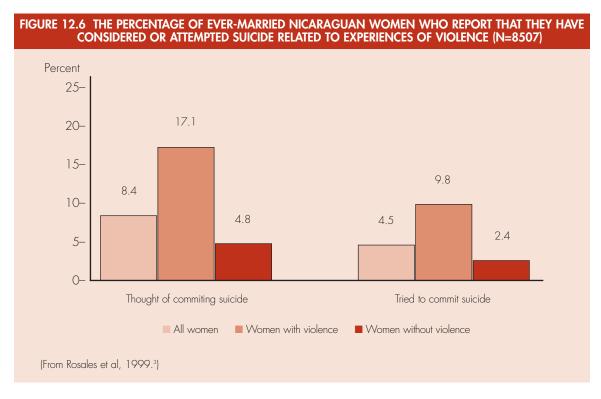
Frequency Once	Moderate Violence % (N) 18 (17)	Severe Violence % (N) 22 (21)	Total % (N) 40 (38)
Occasional (3–5 times)	10 (10)	28 (27)	38 (37)
Frequent (6–20 times)	2 (2)	9 (9)	11 (11)
Very frequent (more than 20 times)	0	11 (11)	11 (11)
Total	30 (29)	70 (68)	100 (97)
(From Ellsberg et al, 1999.4)			

be critically important later on for looking at the relationship between violence and specific health outcomes. (For an example see the association between violence and emotional distress shown in Figure 12.11.)

LOOKING AT ASSOCIATIONS **BETWEEN VIOLENCE AND** OTHER VARIABLES

After completing the descriptive analysis, the next step is to perform cross-tabulations to look for associations between



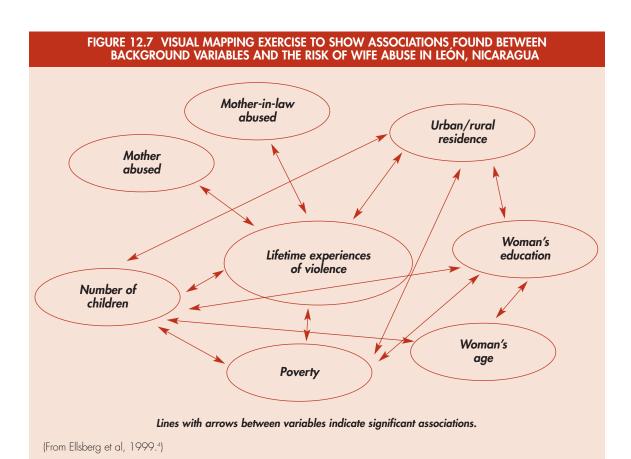


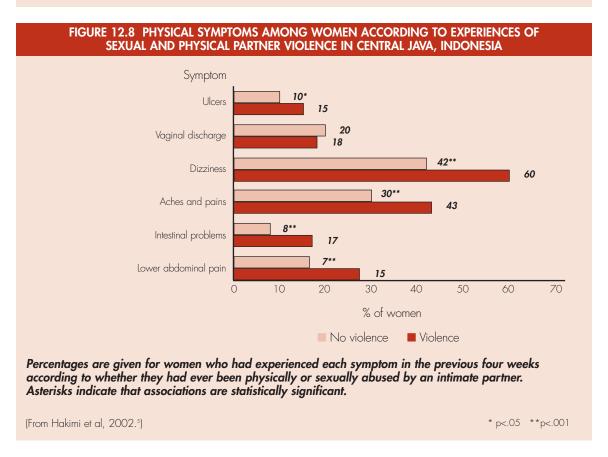
violence and possible risk or protective factors or specific health outcomes. Simply stated, this involves comparing different groups of women to find out whether there is more violence among women with certain characteristics, or whether women who have experienced violence have a greater frequency of certain kinds of problems. Figures 12.5 and 12.6 show how these differences might be presented by means of a bar graph. You will certainly want to look at the associations between experience of violence and such background factors as age, urban rural residence, socio-economic status, education, and number of live children. You should also look at how each of these variables is related to the others. It may be helpful to map the associations out visually, as shown in Figure 12.7.

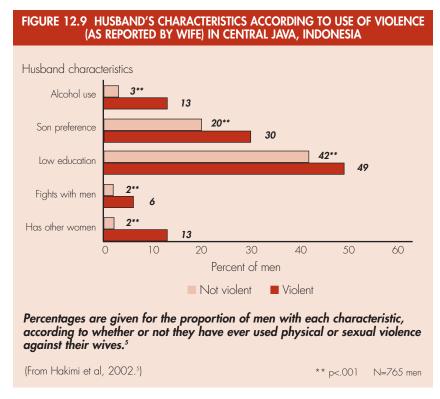
In order to determine whether the prevalence of violence actually varies according to different characteristics such as age or education, you will need to perform tests for **statistical significance**. When used by researchers, the word significance refers not to the importance or size of the difference, but to the likelihood that the associations are real and not simply due to chance. Two of the most common tests for statistical significance are chi-square (X2) and the Student's T test, and most computer packages can perform them automatically. Each test is appropriate for different kinds of data. Therefore, you should consult a statistician to determine the most appropriate statistical tests for your data.

For example, initial analysis of survey data in León, Nicaragua, found that lifetime experience of physical violence was significantly associated with poverty, living in the urban area, and having a mother or mother-in-law who was also abused. No associations were found between violence and a woman's age or education. However, both age and low education were associated with a high number of children, and low education was associated with poverty and living in the rural area (Figure 12.7). In Figures 12.8 and 12.9, bar graphs show findings from a study in Central Java, Indonesia. Figure 12.8 shows that









Indonesian women who had ever experienced partner violence reported more recent symptoms of ill health such as pain, dizziness, ulcers, and intestinal problems. Figure 12.9 shows that Indonesian men who have been violent towards their wives are also more likely to have used alcohol, and to have had extra-marital relationships. They are more likely to have been involved in fights with other men, to prefer sons over daughters, and to have little or no education.

ASSESSING THE VALIDITY OF SURVEY RESULTS

Once you have found what seem to be the most important results from your data, and you have performed basic statistical tests between variables (for example, violence and ill health), you need to assess their validity. This means you need to determine to what degree the study measured what it was supposed to, and whether the findings mean what they are supposed to. Internal validity refers to the extent to

which variations in a specific outcome (for example, the risk of violence) may be attributed to variations in an independent variable (for example women's age or education). External validity refers to the degree to which results from a given study may be used to draw conclusions about a larger population. If a study is performed on a randomly selected population, it should be possible to generalize the results of the study to the general population from which the sample was drawn.

Another important question is Are the findings consistent? That is, do they make sense, according to what is known about the subject locally and internationally? If they differ greatly from previous findings, are there any additional data to support the new results? Are there any aspects of data collection, sampling, design, or analysis that might have altered the results by introducing bias? Has the analysis taken into account possible sources of confounding? The following pages present different ways to address these issues.

The effects of confounding in data analysis

In studying the association between risk factors and a specific problem, confounding can occur when another characteristic exists in the study population and is associated with both the problem and the risk factor under study.

Confounding can have a very important effect on study results, and can create the appearance of a cause-effect relationship that in reality does not exist. Age and social class are often confounders in epidemiological studies. In the study of risk factors for violence, confounding variables can give misleading impressions about what risk factors influence the occurrence of violence.

Stratified analysis and multivariate analysis are two ways to control for the effects of confounding variables.



Stratification involves analyzing data separately using defined categories of the confounding factor, such as age groups. For example, some studies using bivariate analysis (analysis using only two variables) have found that pregnant women were more likely to be abused than non-pregnant women. However, after analyzing the same data stratified by age groups, it turned out that this association was confounded by age. It turned out that being young was the real risk factor for violence rather than pregnancy. It just happened that younger women were more likely to be pregnant than older women. This explained the increased prevalence of violence among pregnant women.

Violence can also be analyzed as a confounding variable for other risk factors, as shown in the following example of a study on mental distress. Preliminary results found that women who had been married at least once in their lives had twice as much emotional distress as women who had never been married (Table 12.3). This would imply that marriage is an important risk factor for mental distress.

However, when the prevalence of mental distress among ever-married women was analyzed separately according to whether women had experienced wife abuse, a large difference was found between the two groups. Thirty-one percent of abused women suffered mental distress, compared to only seven percent of women who had never been abused, which is even less than the prevalence of distress among nevermarried women (Table 12.4).

Since wife abuse is associated with marriage (by definition only ever-partnered women can experience wife abuse) and it is also associated with mental distress, it has a confounding effect on the association between marriage and mental distress (Figure 12.10). Therefore, after stratified analysis it becomes evident that it is wife

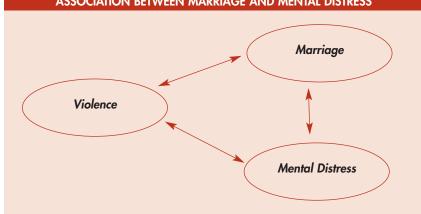
TABLE 12.3 PREVALENCE OF EMOTIONAL DISTRESS ACCORDING TO MARITAL STATUS AMONG NICARAGUAN WOMEN

Selection of women	Percentage of emotional distress
All women 15-49 (n=488)	17%
Ever-married women (n=360)	20%
Never-married women (n=128)	10%
(From Ellsberg et al, 1999.°)	

TABLE 12.4 PREVALENCE OF EMOTIONAL DISTRESS AMONG EVER-MARRIED NICARAGUAN WOMEN ACCORDING TO EXPERIENCES OF WIFE ABUSE

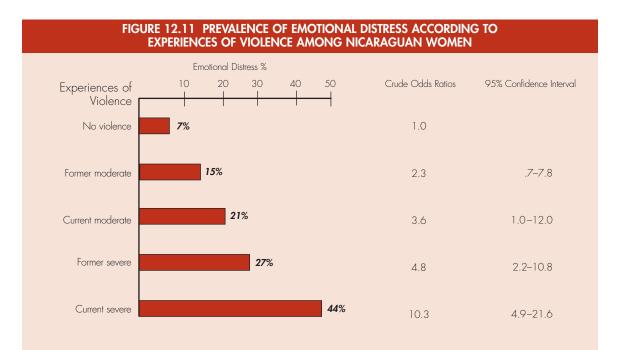
Experience of wife abuse	Prevalence of emotional distress
Never abused (n=172)	7%
Ever abused (n=188)	31%
(From Ellsberg et al, 1999.°)	

FIGURE 12.10 THE CONFOUNDING EFFECT OF VIOLENCE ON THE ASSOCIATION BETWEEN MARRIAGE AND MENTAL DISTRESS



abuse and not marriage itself that accounts for the increase in mental distress among married women.

This analysis is further strengthened by comparing women's current mental distress according to the severity of violence they experienced and when it took place, as shown in Figure 12.11. Breaking down the analysis this way demonstrates that women who were severely abused in the last 12 months were over ten times more likely to be distressed than women who had never been abused. Further, it revealed that the severity of abuse was more important than when it took place, since women experiencing severe



Percentages are given for the proportion of ever-married women who experienced emotional distress in the four weeks prior to the survey, according to whether they had experienced physical partner violence. Violence was classified by severity and by whether it took place within the 12 months previous to the study, or earlier. In the right hand columns, crude (unadjusted) odds ratios and their corresponding confidence intervals are given. Intervals where the lower and upper figures do not include 1.0 are considered statistically significant. (In this case, all types of violence except for former moderate violence are significantly associated with emotional distress.)

violence formerly were still more likely to be currently distressed than women who had suffered only minor abuse, even though it took place more recently.

The use of multivariate analysis to adjust for confounding factors

When it appears that there are several variables confounding an association, then it is no longer practical to use stratified analysis, as it would be excessively complex to perform. For example, in Figure 12.7, we saw that in the León study there were several variables, such as poverty, living in the urban area, and number of children, which were associated with the risk of violence. It was further shown that these three variables are associated with each other as well as with women's level of education. Could it be that poverty is the true underlying factor influencing women's risk of violence, and therefore

that urban women and women with many children are found to have greater levels of violence, simply because they are more likely to be poor? How can we unravel the complex relationships between these variables?

Multivariate analysis techniques, such as logistic regression modeling, are useful for examining the relationships between several explanatory factors and a specific outcome variable. Logistic **regression** helps to uncover the degree to which several explanatory variables are related and to control for confounding variables. In Table 12.5 (next page), the same relationships presented in Figure 12.7 are examined using crude or unadjusted odds ratios as well as multivariate or adjusted odds ratios. The 95 percent confidence intervals are used to assess the statistical significance of the association by indicating that there is a 95



percent probability that the true figure lies between this range. If the range between the lower and upper figure in the confidence interval does not include one, then it can be said that there is a 95 percent probability that the association is not due to chance.

When comparing the crude and multivariate odds ratios for each variable, one can see that they do not vary much in most of the cases. The association between violence and poverty, having more than four children, and a history of family violence in the husband's family are all maintained. Living in the urban area, which had a confidence interval slightly below one in the crude analysis, becomes significant in the multivariate model, while a history of family violence in the wife's family becomes insignificant. After performing the multivariate analysis, it is possible to say that although poverty, urban/rural residence, and high parity are all related, their effect on women's risk of violence is independent and should not be interpreted as the result of confounding.

(From Ellsberg et al, 2000.8)

TABLE 12.5 ASSOCIATIONS BETWEEN BACKGROUND FACTORS AND PREVALENCE OF VIOLENCE AMONG 360 EVER-MARRIED NICARAGUAN WOMEN AGES 15–49

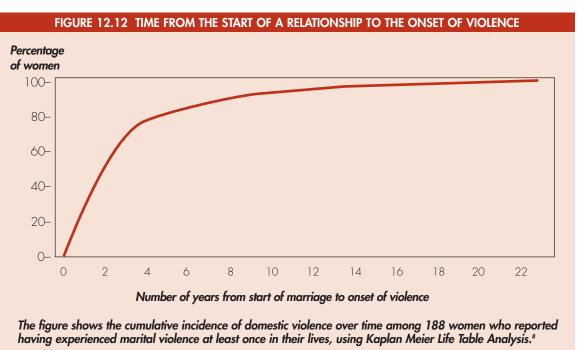
Variable Poverty	Categories Nonpoor Poor	Crude OR (95% CI) 1.0 1.91 (1.12–3.23)	Adjusted OR (95% CI) 1.0 1.82 (1.03–3.23)
Zone	Rural Urban	1.0 1.62 (.94–2.78)	1.0 2.07 (1.12–3.82)
Number of children	0-1 2-3 4 or more	1.0 1.40 (.82–2.39) 2.77 (1.59–4.82)	1.0 1.34 (.74–2.43) 2.23 (1.21–4.15)
Family history of abuse	No history in wife's family Wife's mother abused	1.0	1.0 1.28 (. <i>7</i> 9–2.09)
	No history in husband's family Husband's mother abused	1.0	1.0

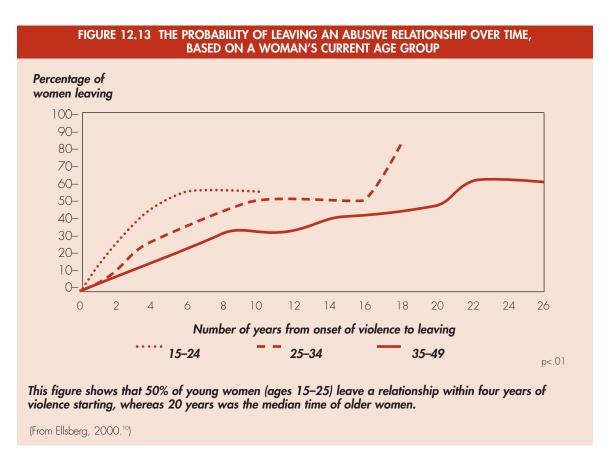
Crude and adjusted odds ratios are given (together with 95 percent confidence intervals) for having experienced violence at least once in their lives.

(From Ellsberg et al, 1999.4)

Using advanced statistical analysis creatively

In earlier sections of this chapter, we presented the most commonly used techniques for statistical analysis of survey data





on violence. However, additional insight may be revealed by the creative use of more advanced statistical techniques.

For example, **life table** or **survival** analysis was used to gain a deeper understanding of the relationship between violence and high parity in Nicaragua. Many international studies have found a similar association.7 One interpretation for this is that having many children places additional stress on a marriage and increases a woman's likelihood of being beaten by her husband.

However, using life table analysis, a statistical technique which measures the probability of events occurring over time, it was possible to determine that violence began early on in relationships, in many cases well before women had started bearing children. Figure 12.12 shows that 50 percent of violence begins within two years of marriage, while 80 percent of abuse starts within four years. This

implies that high parity, instead of being a risk factor for abuse, is more likely to be a result of violence, because battered women are less likely to be able to control the timing of sex or the use of birth control.

The same techniques were applied to the likelihood of a woman leaving an abusive relationship, and it was found that 70 percent of women eventually did leave their abusers, although some women stayed as long as 25 years or more before separating. Stratifying this analysis according to age groups shows that younger women are more likely to have left an abusive relationship within four years, compared to women between 35-49 years (Figure 12.13). This indicates that younger women are less likely to tolerate abuse than older women. In order to use survival analysis techniques, it is necessary to collect detailed data regarding each of a



woman's relationships: when did it start, how long did it last, was there violence, and if so, when did the first and last incidents of violence take place. These types of analyses are somewhat complicated to perform and interpret, so it is important to consult with an experienced statistician.

INTERPRETING THE RESULTS

The process of data analysis will often take longer than you initially expect. However, you can plan data analysis in stages, so that initial findings, such as prevalence and descriptive characteristics, can be made available as soon as possible to the communities and local institutions that have been supporting the research, and that will be anxiously awaiting results. Further analysis can be performed over a longer period to explore some of the more interesting findings in greater depth. Box 12.2 presents guidelines for writing up research results for publication in scientific journals. Chapter 14 will discuss in detail how research results may be tailored to fit the needs of different groups.

When interpreting and writing up the results of data analysis, it is important to be cautious. Each research design yields different kinds of data, with their respective limitations. Be careful not to draw conclusions that are not supported by the data, as overstating your results can seriously undermine the credibility of the research. People are more likely to listen to your findings when you are open about whatever limitations the study had in terms of design, data collection, or analysis. Some examples of common pitfalls are the following:

■ Inferring causal relationships from cross-sectional data. Cross-sectional surveys can highlight associations between

BOX 12.2 SUGGESTED GUIDELINES FOR WRITING A SCIENTIFIC PAPER

Abstract

Approximately 100 words.

Background

Literature, national context, objectives.

Describe the study population, how the sample was selected, what instruments were used, how the fieldwork was conducted, how data were analyzed, how ethical clearance was obtained, and any special measures, such as safety procedures.

Results

This section should describe all the major results of data analysis, including relevant tables and figures, and measures of statistical significance.

Discussion

The purpose of this section is to interpret the meaning of data, assessing the validity and generalizability, possible sources of bias, how the findings relate to international and national studies on the same subject, and possible explanations for the most important findings.

Conclusions

These are sometimes included in the discussion section. How might these findings be used for improving interventions and policy? What are areas that might benefit from future research?

References

Make sure to include citations from the most relevant literature in the field of study.

(From Persson and Wall, 2003.1)

two variables, but unless you have good information about when different conditions or events occurred it is difficult to know with certainty what came first. A good example of how causal relationships can be misinterpreted is the relationship between parity and violence presented in the last section. It is a good idea when presenting results from cross-sectional surveys to talk about "associations" rather than causes. The discussion section can assess which variables are most likely to be causes or outcomes, based on your conceptual framework and other studies on the subject.

Inferring causal relationships from bivariate analysis. As we showed in the

- example on marriage and emotional distress, other variables may confound a relationship between two variables. If you have not performed stratified or multivariate analysis, it is wise to be cautious in interpreting your results.
- Generalizing conclusions for different populations than the study population. Results that are representative for one region are not necessarily true for other regions in the country, or for the country as a whole. This does not mean that regional studies cannot provide important insights that are relevant for a much broader context. There are many examples of regional studies that made critical contributions for guiding national policies and programs. However, it is still important to be careful in stating clearly what the limitations of the sample are, both in terms of its power to capture important associations and its generalizability.
- 1. Persson LÅ, Wall S. Epidemiology for Public Health. Umeå, Sweden: Umeå International School of Public Health; 2003.
- 2. Yoshihama M, Sorenson SB. Physical, sexual, and emotional abuse by male intimates: Experiences of women in Japan. Violence and Victims. 1994;9(1):63-77.
- 3. Rosales J, Loaiza E, Primante D, et al. Encuesta Nicaraguense de Demografia y Salud, 1998. Managua, Nicaragua: Instituto Nacional de Estadisticas y Censos, INEC; 1999.
- 4. Ellsberg MC, Peña R, Herrera A, Liljestrand J, Winkvist A. Wife abuse among women of childbearing age in Nicaragua. American Journal of Public Health. 1999;89(2):241-244.
- 5. Hakimi M, Nur Hayati E, Ellsberg M, Winkvist A. Silence for the Sake of Harmony: Domestic Violence and Health in Central Java, Indonesia. Yogyakarta, Indonesia: Gadjah Mada University; 2002.
- 6. Ellsberg M, Caldera T, Herrera A, Winkvist A, Kullgren G. Domestic violence and emotional distress among Nicaraguan women: Results from a population-based study. American Psychologist. 1999;54(1):30-36.
- 7. Kishor S, Johnson K. Domestic Violence in Nine Developing Countries: A Comparative Study. Calverton, MD: Macro International; 2004.
- 8. Ellsberg M, Peña R, Herrera A, Liljestrand J, Winkvist A. Candies in hell: Women's experiences of violence in Nicaragua. Social Science and Medicine. 2000;51(11):1595-1610.
- 9. Ellsberg MC, Winkvist A, Peña R, Stenlund H. Women's strategic responses to violence in Nicaragua. Journal of Epidemiology and Community Health. 2001;55(8):547-555.
- 10. Ellsberg M. Candies in Hell: Research and Action on Domestic Violence in Nicaragua [Doctoral Dissertation]. Umeå, Sweden: Umeå University; 2000.