

Sample-Size Requirements for Clinical Trials of Safety-Engineered Sharp Devices

By Gina Pugliese, R.N., M.S., Teresa P. Germanson, Ph.D., Judene Bartley, M.S., M.P.H.,
Judith Luca, R.N., Lois Lamerato, Ph.D., Jack Cox, M.D., and Janine Jagger, M.P.H., Ph.D.

*This article originally appeared in *Infection Control and Hospital Epidemiology* (2001;22:456-458) and is reprinted, in abridged form, with permission.*

THE NEEDLESTICK SAFETY AND Prevention Act of November 2000, which required healthcare employers in the United States to purchase safety-engineered needles and sharp medical devices, greatly accelerated the transition to safety technology. Today, the majority of needles and sharp devices used in clinical settings in the U.S. are of the safety variety, and the number of conventional devices in use has been declining. There remains a continued need, however, for product evaluations in order to assess the performance of safety devices.

Evaluations can be conducted in a number of ways. The most common is the informal device evaluation or product trial, in which subjective feedback from users is elicited; this type of evaluation has no minimum requirement for the number of devices that must be evaluated. Informal product trials are usually brief and involve clinical observations of a relatively small number of devices. They can provide valuable information about user preferences and product characteristics when healthcare institutions are considering the adoption of new devices.

However, such informal evaluations cannot be used to draw objective conclusions about user injury rates or the safety performance of specific devices. This can only be accomplished through a properly

designed, large-scale study, with the results subjected to statistical analysis. The table on page 71 shows the sample sizes of conventional and safety devices required in order to make statistically valid comparisons of injury rates between the two groups.

Safety Efficacy Studies

Safety efficacy studies are based on comparisons of device-specific injury rates—in this case, rates for conventional devices and their safety counterparts. Because needlesticks are, in statistical terms, rare events, it can be difficult to achieve a sample size large enough to yield statistically valid information. According to published studies, needlesticks occur in the range of 1 to 37 injuries per 100,000 devices used.¹⁻⁶ There is an inverse relationship between injury rates and required sample size: the lower the injury rate, the larger the denominator (i.e., number of devices) must be in order to attain an acceptable statistical power of 80% or 90%. Statistical power is the ability of a test to show a statistically significant difference between groups, if a true difference exists.

Sample Size Calculations

To determine the sample size needed to achieve a given level of statistical power, the researcher must first decide the minimum reduction in injuries from the safety device that would be clinically meaningful for that particular study. For instance, should the device prevent at least 75% of injuries, or would it be acceptable to prevent as few as 25%? Reasonable expectations for injury

reductions from safety-engineered sharp devices can be derived from the literature; published studies show reductions from safety devices ranging from a low of 25%³ to a high of 89%.⁵ It is up to the researcher to decide what is acceptable under the specific circumstances of the trial. This decision, in turn, determines the number of devices that must be used in the study to achieve statistical significance. It is worth noting that a 100% reduction in injuries is an unrealistic expectation, unless the safety device completely eliminates the sharp.

The table on page 71 correlates sample-size requirements for the devices being compared (safety and conventional) with the projected injury rates for the conventional device and the desired level of injury reduction for the safety device. The sample sizes given in the table are designed to achieve a statistical power of 80%, where alpha is set to 0.05 in two-tailed testing, and are calculated using Fisher's exact test.⁷ Since the distribution of rare-event data may not meet the normality assumptions of Pearson's chi-square test, an exact test is the preferred analytic procedure. The selected parameters are derived from current literature and should provide realistic assumptions for determining sample sizes.

For example, if the projected baseline injury rate for the conventional device is 15 injuries per 100,000 devices used, and the researcher hopes to demonstrate a 75% reduction in injuries with the safety device, then 125,000 conventional devices and 125,000 safety devices

would need to be used to achieve a statistical power of 80%. On the other hand, if the baseline injury rate for the conventional device is 5 per 100,000 and the researcher wants to show the same 75% reduction in injuries, the sample-size requirement is 375,000 for each kind of device. If the investigator establishes 25% as the minimum reduction worth detecting, then the sample-size requirement would be 1.15 million per group, given a baseline injury rate of 20 per 100,000.

Discussion

The calculations presented in the table show that the number of devices required to compare needlestick rates for conventional and safety devices is enormous—even for sample sizes on the lower end of the scale. At the high end, more than four million devices per group would be needed to achieve statistical significance if a conventional device causing 5 injuries per

100,000 devices were compared to a safety device reducing injury rates by 25%. At the low end, 94,000 devices per group would be needed to achieve statistical significance if a conventional device causing 20 injuries per 100,000 devices were compared to a safety device that reduced injury rates by 75%.

For the majority of healthcare facilities, these sample-size requirements would exceed annual device usage in most device categories. In the largest hospitals, an estimated 1.5 to 3 million syringes are used per year; for intravenous (IV) catheters, usage ranges from 75,000 to 300,000 devices per year, with a similar range for winged steel needles and phlebotomy needles. In average-sized or

small hospitals, the number of devices used annually in these categories is much lower.

The investigator contemplating a single-center trial must first determine if the annual usage of the device to be evaluated is adequate to meet minimum sample-size requirements. If not, a multi-center study will likely be necessary. Another alternative is to extend the trial for the amount of time necessary to achieve an adequate sample size—if the results would still be worth knowing by the time

statistically significant.

Many challenges face investigators as they embark on efficacy trials of safety-engineered devices, but none is more important than assuring that the sample size is adequate to meet statistical requirements. This will become even more challenging as safety-engineered devices become the predominant technology in the medical device marketplace. Eventually, safety devices with very low baseline injury rates will be tested against other, similar safety devices to determine which is “safest.” The result is that sample-size requirements for such studies may reach prohibitive extremes, making multi-center collaborative studies a necessity in order to meet statistical demands. □

References

1. Jagger J, Hunt EH, Brand-Elnaggar J, Pearson RD. Rates of needle-stick injury caused by various devices in a university hospital. *N Engl J Med* 1988; 319(5):284-288.
2. Ippolito G, De Carli G, Puro V, Petrosillo N, Arici C, Bertucci R, Jagger J, et al. Device-specific risk of needlestick injury in Italian healthcare workers. *JAMA* 1994;

Table 1. Sample-size requirements for detecting reductions in needlestick rates†

Injuries per 100,000 conventional needles	% reduction with safety devices	Number of devices required per device type
5	25%	4,600,000
	50%	1,000,000
	75%	375,000
10	25%	2,300,000
	50%	500,000
	75%	186,000
15	25%	1,540,000
	50%	340,000
	75%	125,000
20	25%	1,150,000
	50%	250,000
	75%	94,000

† Fisher's exact test, alpha=0.05, power=80%, two-tailed testing.

the trial ended.

Undertaking a study without an adequate sample size increases the chance that statistical significance will not be achieved—even if the safety device is effective. Consider, for example, a study in which 150,000 needles are used in each group, conventional and safety, and the conventional device has a true baseline injury rate of 5 injuries per 100,000 needles. If the true relative effectiveness of the safety device is 50%, then the probability of obtaining statistically significant results in two-tailed testing, where alpha=0.05, is only 27%. In other words, it is most likely that the results from the study will not be

272(8):607-610.

3. Centers for Disease Control and Prevention. Evaluation of safety devices for preventing percutaneous injuries among health-care workers during phlebotomy procedures—Minneapolis-St. Paul, New York City, and San Francisco, 1993-1995. *MMWR* 1997; 46(2):21-25.
4. Jagger J, Bentley MB. Injuries from vascular access devices: high risk and preventable. Collaborative EPINet Surveillance Group. *J Intraven Nurs* 1997; 20(6 Suppl):S33-S39.
5. Mendelson MH, Chen LBY, Finkelstein LE, Bailey E, Kogan G. Evaluation of a safety IV catheter (Insyte Autoguard, Becton Dickinson) using the Centers for Disease Control and Prevention (CDC) National Surveillance System for Hospital Healthcare Workers database. *Infect Control Hosp Epidemiol* 2000;21:111. Abstract.
6. Chen LBY, Bailey E, Kogan G, Finkelstein LE, Mendelson MH. Prevention of needlestick injuries in healthcare workers: 27 month experience with a resheathable safety winged steel needle using CDC NaSH Database. *Infect Control Hosp Epidemiol* 2000;21:108. Abstract.
7. Fisher, RA. *The Design of Experiments*. Edinburgh, Scotland: Oliver and Boyd; 1935.