

Use of Scaled Venn Diagram (SVD) to Present Epidemiological Concepts

Larssen, R.B. Rolf Bjerke Larssen, Norwegian School of Veterinary Science, POB 8146 Dep, N-0033 Oslo, Norway, rolf.b.larssen@vetsh.no

Summary

This presentation demonstrates how a Scaled Venn Diagram (SVD) can be used to visualize epidemiological concepts such as independence, population attributable risk, and the effect of prevalence on positive predictive value.

Introduction

Teaching medical statistics and epidemiology is a pedagogic challenge. One strategy is to visualize concepts and phenomena as much as possible. Then one tool is the ordinary Venn diagram, which is useful to show relative abundance of categories, and to show to which extent categories overlap. But it is not very useful for demonstration of relationships between variables e.g. (in)dependence.

To overcome this shortcoming of the traditional Venn Diagram, I have during the last 20 years of teaching used what could be called a Structured or Scaled Venn Diagram (SVD) to illustrate:

- General concept of independence between variables.
- Measures of association and effect – e.g. the population attributable risk.
- Diagnostic tests – e.g. effect of prevalence on positive predictive value.
- Systematic errors – e.g. confounding and interaction.

In spite of the usefulness of the SVD it seems not to be used to any extent in epidemiology textbooks or other teaching material, thus this promotion at ISVEE X.

Materials and methods

Epidemiologists are known to “think in 2x2 tables”, and the SVD can be regarded as a visualization aid for communicating this thinking. Thus the starting point is the 2x2 table with its typical layout of “Disease +/-“ as columns and “Factor +/-“ as rows, and the letters a, b, c and d denoting the 4 cells. The procedure for making the corresponding SVD is:

1. Draw a rectangle.
2. Draw a horizontal line (if Cohort study) or a vertical line (if Case control study) or either (if Cross sectional study) dividing the rectangle in two parts proportional to group size.
3. Within each of the groups demarked above, draw a line perpendicular to the first line. dividing the subgroups within the previous groups.

Each of the 4 sides of the rectangle can have a meaningful use as scale for proportion going from 0.0 to 1.0, but the relative proportions between all the 4 cells will be representative for the study population only if we have a cross sectional study sample. However if there is independence between the two variables in the corresponding 2x2 table, this will always show up as a perfect “thread cross” between the horizontal and the vertical dividing lines as shown in Figure 1.

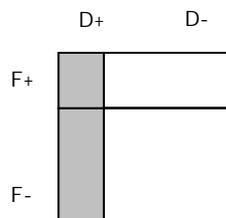


Figure 1. Independence in a 2x2 table – the “thread cross” situation

Table 1. Example data and definition of symbols used in the 2x2 table. Prevalence of intervertebral calcification in Dwarf dachshund (F+) vs Standard dachshund. Stigen O. (1991).

Exposed (Dwarf) F+	37 a	65 b	102	36.3%
NonExp (Standard) F-	50 c	173 d	223	22.4%
Sum	87	238	325 n	26.8%

The example data (Table 1) are from a cross sectional study of back problems in Dachshund, and are presented as SVD in Figure 2. Several measures of association and effect can be calculated. Some express the excess risk for the exposed group (Attributable risk=14.1%, Proportional attributable risk=39.0%, Relative risk=1.6); while others express the excess risk for the population (Population attributable risk=4.3%, Population proportional attributable risk=16.2% (PAF), Population relative risk=1.2). However, all these measures are just different ways of presenting the extra 14% of disease among the dwarfs, i.e. the extra 14 dwarfs in our data that have the disease due to being dwarfs as shown in the SVD to the right in Figure 2+3.

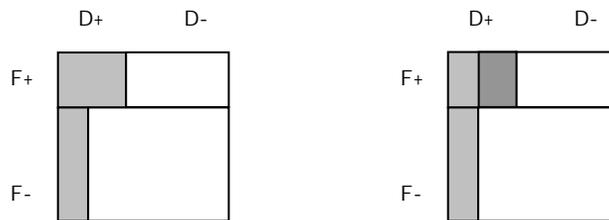


Figure 2. SVD demonstrating the importance for the exposed group. I.e. showing the proportion of diseased among dwarfs (F+) vs among standard dachshund (F-). The excess cases among dwarfs show up as the darkest area in the SVD to the right.

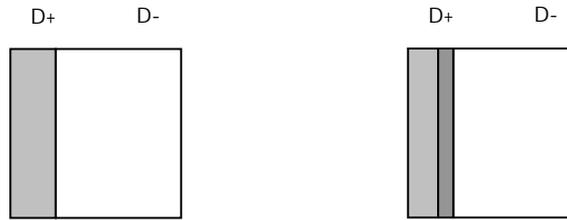


Figure 3. SVD demonstrating the importance for the population showing the proportion of diseased in the population disregarding the subdivision in dwarfs (F+) and standard dachshund (F-). The excess cases in the population show up as the darkest area in the figure to the right.

Properties of diagnostic tests is an important epidemiological subject, and one important point in teaching is to make people understand how the prevalence affects the positive predictive value, $PPV = a/(a+b)$. Figure 4 shows how SVD can be used to illustrate how PPV goes down when prevalence goes down in a diagnostic test situation.

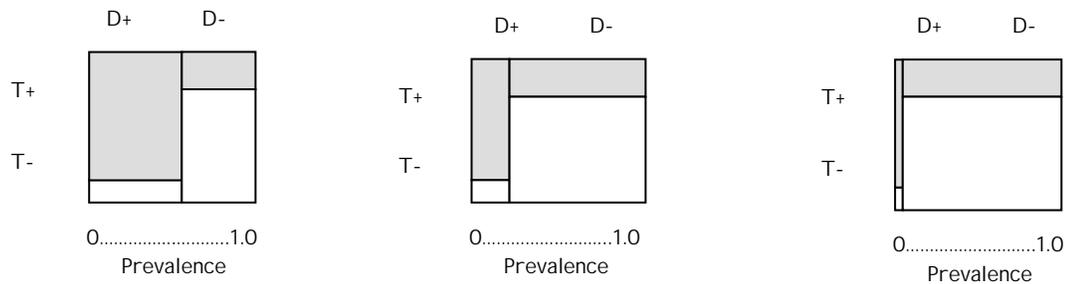


Figure 4. Use of SVD to demonstrate the effect of prevalence on positive predictive value (PPV) of a diagnostic test.

Discussion

In addition to its use related to the 2x2 table, the SVD can also be used to illustrate multivariate situations as described by Marshall, R.J. (2001) and by Eide, G.E., Heuch, I. (2001). However, its greatest potential is probably as a teaching aid to help newcomers in the field of epidemiology to differ between basic concepts. It may also be beneficial if this way of visualization is adopted by statistical software packages.

References

1. Marshall, R.J. (2001) Displaying clinical data relationships using scaled rectangle diagrams. *Stat Med*, 20,1077-88
2. Eide, G.E., Heuch, I. (2001) Attributable fractions: fundamental concepts and their visualization. *Stat Methods Med Res*, 10, 159-93
3. Stigen, O. (1991) Calcification of intervertebral discs in the dachshund. A radiographic study of 327 young dogs. *Acta Vet Scand*, 32, 197-203.