Interpretation and visualisation of data from dairy herds

Kristof Hermans, Geert Opsomer, Willem Waegeman, Sander Moerman, Jenne De Koster, Mieke Van Eetvelde, Bonifacius Van Ranst, Miel Hostens

Vets often use key performance indicators (KPIs) to evaluate the actual performance status of a dairy herd. Basic knowledge of data analysis is necessary to interpret these KPIs correctly, but unfortunately vets often lack the training and knowledge required to do this. This article aims to make vets aware of common data interpretation errors and discusses four data misinterpretation pitfalls – variation, momentum, lag and bias – as well as introduces the concept of data visualisation, such as the use of graphs and charts, to help vets avoid these pitfalls.

Kristof Hermans qualified from the University of Ghent in 2010, where he is currently undertaking a PhD on herd health management.

Geert Opsomer qualified from the University of Ghent in 1989 , where he is currently an associate professor.

Willem Waegeman qualified from the University of Ghent in 2004 with an MSc in computer science. He holds a PhD in engineering and is currently working as a professor in the Department of Mathematical Modelling, Statistics and Bioinformatics, University of Ghent.

Sander Moerman qualified from the University of Ghent in 2014. He currently works in a large animal practice in Roeselare, Belgium.

Jenne De Koster qualified from the University of Ghent in 2009, and obtained a PhD in veterinary medicine in 2016. He currently works as an assistant on herd health management at the University of Ghent.

Mieke Van Eetvelde qualified from the University of Ghent in 2009, where she currently works as an assistant on herd health management.

Bonifacius Van Ranst qualified from the University of Ghent in 1980. Since 1987 he has been working as a farm animal practitioner and dairy consultant.

Miel Hostens qualified from the University of Ghent in 2008 and obtained a PhD in veterinary medicine in 2013. He currently works at the university as a Postdoctoral assistant on herd health management.

doi: 10.1136/inp.k2166

OVER the past few decades, the dairy industry has changed from small family-owned businesses to large facilities with hundreds, or even thousands of animals. Simultaneously, technological development has enabled the collection of high volumes of dairy data at a tremendous rate. The ability to gather data has outpaced the speed of data analysis. As such, this has altered the way data is presented to the veterinarian. Raw data is preprocessed using, for example, herd health or managing software, in calculations, algorithms and visualisation tools, such as graphs and charts, to ease the life of the practitioner. Most often, practising vets do not want to process the data themselves, but rather prefer to receive a quick and easy-to-interpret report of the actual performance status of the dairy herd.

Herd performance information can be provided through the aggregation of data into key performance indicators (KPIs). To interpret KPIs correctly and overcome pitfalls in data interpretation, vets must have knowledge on what the KPI means, how it is calculated and how to interpret the result, as well as a basic knowledge of data analysis. Unfortunately, vets often lack training and the knowledge to be able to analyse data thoroughly (Mee and Buckley 2003).

This article aims to make vets familiar with common data interpretation and visualisation rules and errors. We don't suggest how to visualise specific dairy herd KPIs, but rather provide the basic principles of data interpretation and visualisation to practitioners.

Features of a key performance indicator

A KPI can be a simple average (eg, age at first insemination) or the result of a complex computation (eg, net profit gain per cow per year). An inherent feature of a KPI is the fact that it can be calculated within a specific dimension. One of the most important dimensions is time. Time allows aggregation and summarisation of data within specific time frames (ie, month, quarter, year). The second most important dimension is the group in which a KPI is calculated. The group may be composed according to a certain parameter: place (region, country), animal group (whole herd, first lactation cows only), or people (the farmer or vet carrying out inseminations). The time and group dimensions determine whether a KPI calculation is cross-sectional or in a cohort (Mann 2003). A cross-sectional calculation measures a KPI at a certain moment in time. A cohort follows a group of animals with the same property (Fig 1).

Group size can strongly influence the reliability of a KPI (De Kruif 1978). If we consider an increase in conception rate from 30 per cent to 35 per cent, in theory, with a power of 80 per cent and a significance level of 5 per cent, 1075 animals should be included in the calculation for this increase to be significant.

Data used for the calculation of a KPI is not always normally distributed, having serious consequences for the interpretation of numbers such as mean, median and mode (Wonnacott and Wonnacott 1972). If we consider the quantity of animals per lactation number as normally distributed, the average lactation number would be higher in comparison with the non-normal distribution (Fig 2).

Interpretation pitfalls

Some major pitfalls arise when it comes to the interpretation of a KPI, especially when using numbers instead of graphs. Four major concepts increase the chances of misinterpreting data: variation, lag, momentum and bias (Overton 2009).

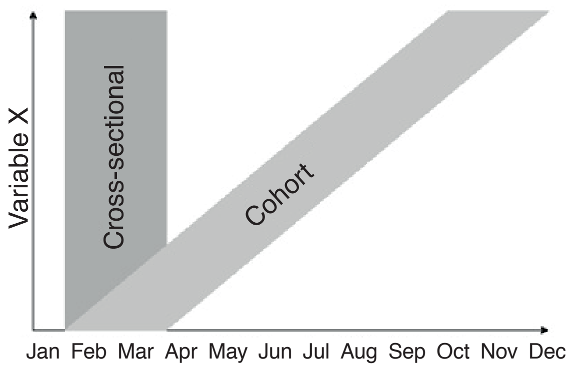

Fig 1: Chart explaining the difference between a cross-sectional and a cohort measurement. A cross-sectional measurement is a snapshot on a certain moment in time. For example the average daily milk production in February and March. A cohort measurement is the opposite: the same measurement is been made over a period of time. For example, the average daily milk production from February to December of cows that have calved in February and March

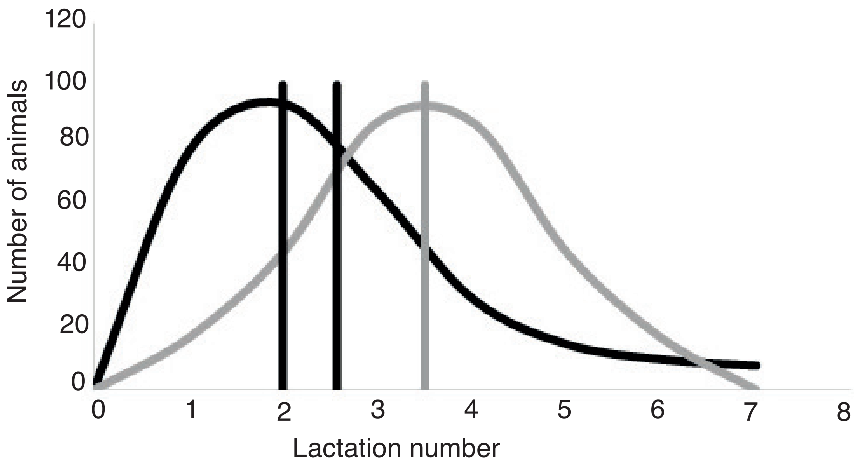

Fig 2: Line graph showing the difference between non-normally (black) and normally (grey) distributed data. We considered a total of 300 animals and plotted them according to their lactation number. The black curve line shows the distribution in reality, the grey curve line shows a virtual normal distribution. The vertical lines show the average (mean and mode) for both distributions. For the normal distribution, the average, the median and the mode are the same (grey line = 3.5). For the non-normal distribution, the average is 2.57 while the median and mode is 2

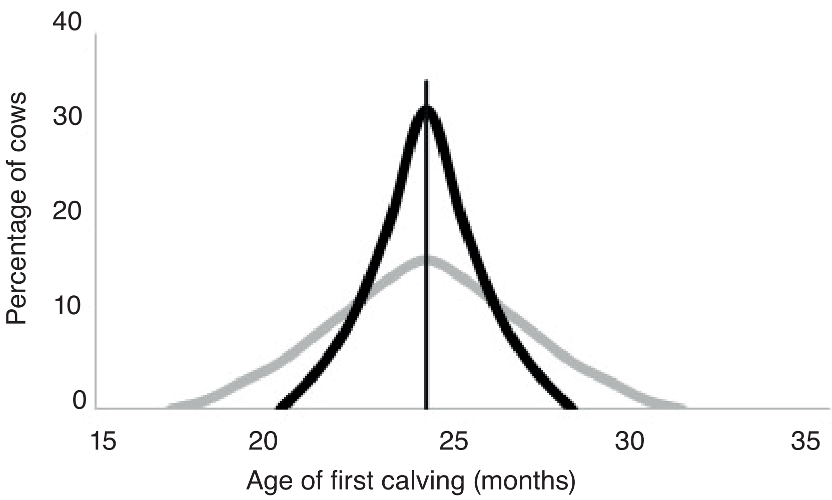

Fig 3: Line graph showing variation in age of first calving of dairy cows in two herds. The age of 24 months is considered as being optimal. For both the grey and the black curves, the average is 24 months; however, the black curve is preferred due to the fact that there is less variation

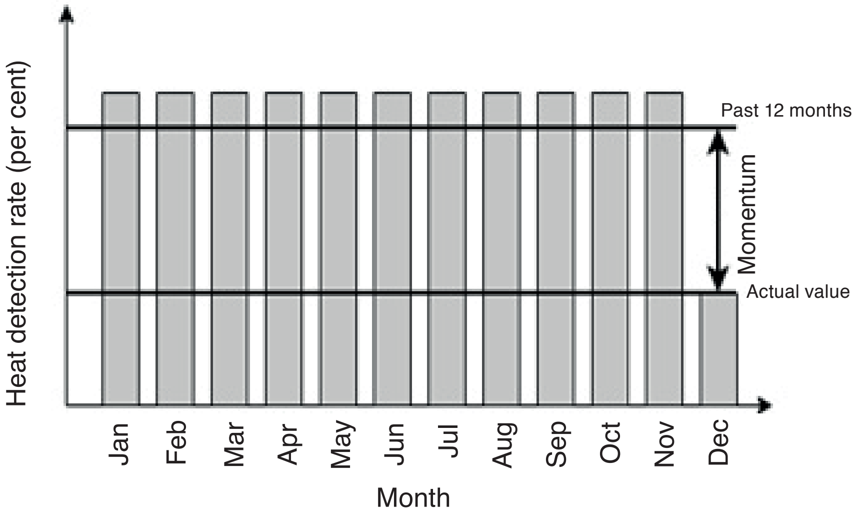

Fig 4: Bar chart demonstrating momentum using heat detection rates (HDR) over a 12-month period. The HDR for December is too low. If we evaluate HDR as the average since January (ie, over the past 12 months), heat detection in December is mistakenly considered as acceptable. The gap between the actual value and the past 12 months average represents the momentum

Variation

All (biological) processes are prone to variability. Variation is known as the deviation from the mean. It is a measure for the scatter of a certain data set. Variation is a threat for correct data interpretation as people are eager to use averages for evaluating and comparing group or herd parameters. Averages usually hide those individual animals that need special attention (van der Leek 2015).

Variation can be visualised using histograms (Fig 3). An excellent example of the importance of variation in dairy management is provided by Bach and Kertz (2010), where they state that an average age at first calving of 26 months can be better than the optimal 24 months if there is less variation across the whole herd; for example, 24 to 28 months instead of 18 to 30 months.

Momentum

Momentum occurs when historic data mask or blur actual performances (Overton 2009). It is known as the responsiveness of averages to recent changes in performances. Momentum can occur in two ways: inferior historical performances can mask current good performances, while excellent historic performances can blur current substandard performances. For example, the average heat detection rate over the past couple of months may not be representative for the rate seen in the most recent month (Fig 4).

A solution to avoid momentum is to evaluate only recent data, although this is not always possible. A larger timeframe is needed when KPIs are calculated for smaller herds or groups in order to have a sufficient number of animals for the calculation (Fetrow and others 1990). For example, if we consider two animals being included in the calculation and one of them is not inseminated, we get a reduction of the insemination risk (IR) of 50 per cent. If we include 10 animals in the calculation and one animal is not inseminated, the reduction in IR is now only 10 per cent.

Lag

Lag is defined as the period between the moment an event occurs and the moment it is measured (Eicker and others 2006). For example, a recent reproductive problem will not have an effect on the historical calving interval during the following nine months (Overton 2009). Conversely, when the historical calving interval suddenly increases, this is likely to be the effect of a reproductive problem that was ongoing nine or more months ago (Fig 5).

Bias

Bias can be caused by the inclusion or exclusion of individuals during the calculation of a KPI (Steineck and Ahlbom 1992). Inclusion or exclusion can be inherent to the calculation. For example, the exclusion of heifers in the historical calving interval, due to the fact that the calculation is based on two consecutive parturitions (Fetrow and others 1990).

Incorrect inclusion or exclusion arises when unsuitable animals or events are included or excluded in calculations (Overton 2009). Including barren animals (or those with 'do not breed' status) in a fertility KPI generally lowers the reproductive performances of the herd, although there may be no reproductive problem. A common cause of this bias is the fact that farmers do not routinely record animals as 'not to breed'. Hence, these animals are considered ready to be served and to get pregnant, and are included in calculations. Logically, bias also occurs when data is missing or incomplete (Nebel and DeJarnette 2007).

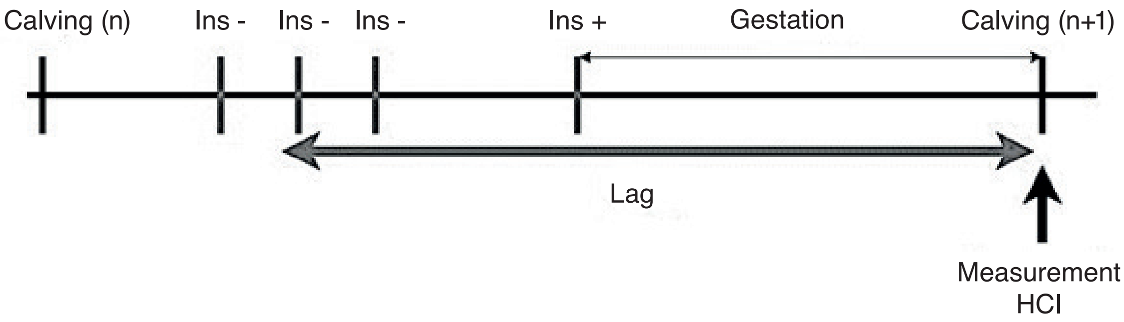

Fig 5: Graph highlighting lag using the historical calving interval (HCI). The HCI can only be measured after two or more parturitions. Gestation is a fixed period of 280 days. The length of the HCI is determined by the period between previous calving and the latest successful insemination (ie, Ins +). This period is situated more than 280 days before calculation of HCI, leading to a lag time of at least 280 days

Visualisation

Visualisation is a useful tool to avoid misinterpretation of numbers. We have reproduced and simplified graphs from Anscombe (1973) (ie, the Anscombe quartet), using milk per day (MPD) in correlation with the days in milk (DIM) for four virtual groups of five cows (Fig 6). In all groups, the average MPD is 20 kg/day and DIM is 50 days. Although the averages are the same, the distributions of the data are completely different, highlighting the importance of good data visualisation.

The human brain is extremely well developed when it comes to recognising visual patterns in data. The Gestalt effect is a theory which describes the ability of the human brain to translate points and lines into whole cohesive forms (Ellis 1938). The Gestalt effect can be summarised as, 'the whole is bigger than the sum of the elements'. The theory includes six major laws.

The first law handles closure, referring to the fact that people tend to ignore gaps and proceed to complete contour lines. The second law regards proximity; people tend to see objects near each other as one big object. Continuity is the third law and involved people grouping objects together as long as they are co-linear or following the same direction. The fourth law – figure and ground – refers to the fact that viewers always perceive an object and a background. Similarity, the fifth law, refers to the fact that objects with the same attributes are likely to be grouped as one. The sixth and most important law is the law of 'Prägnanz': the brain loves simple elements instead of complicated shapes; therefore, this law is also called the law of simplicity.

However, application of these laws can lead to graphs and charts being misinterpreted.

Misrepresentation of data in graphs and charts, whether on purpose or not, is called deceptive visualisation. Deceptive visualisation is categorised into message reversal and message exaggeration or understatement. Message reversal shows the data incorrectly (ie, yes becomes no), while message exaggeration refers to the fact that the data are represented in incorrect proportions.

How to enhance data visualisation

Each graph is composed of several elements such as points, lines, axes and grid lines. Those elements have properties such as colour, size and thickness. The

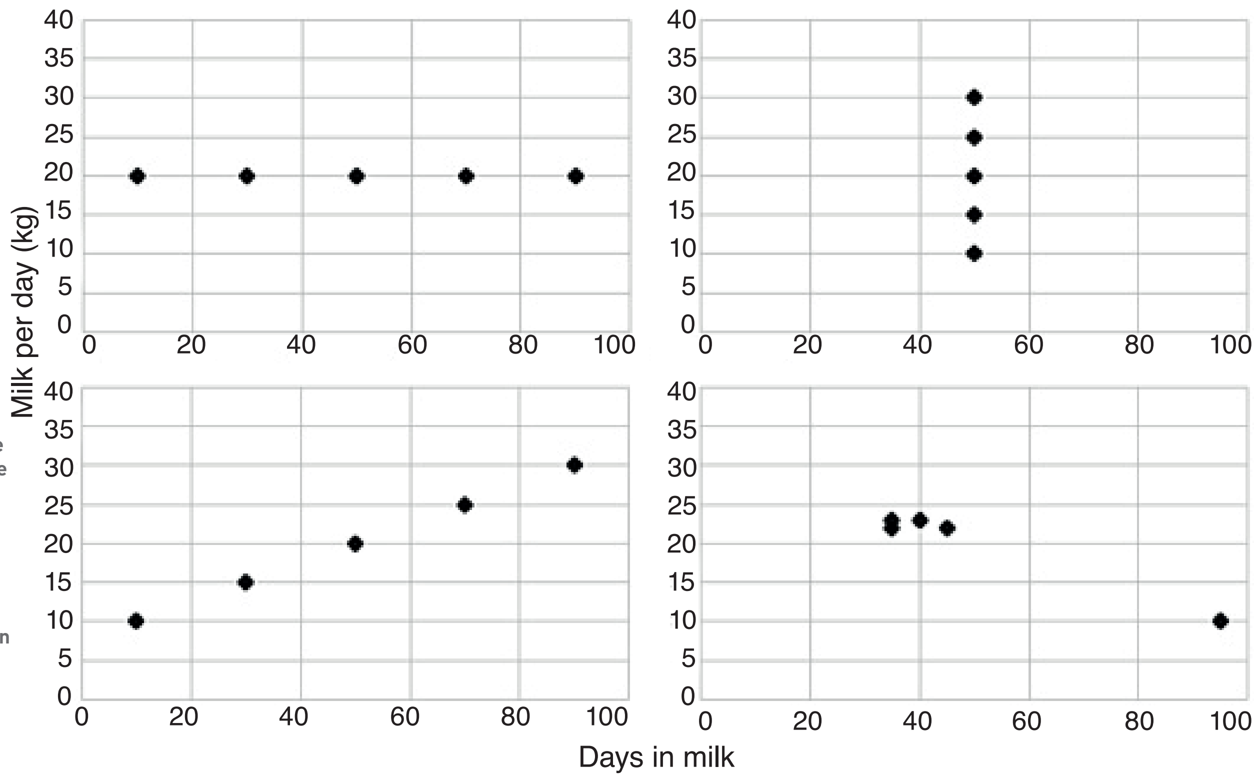

Fig 6: Each of the four graphs contains data of five individual cows. The average amount of milk per day (20 kg) and days in milk (50 days) for each group is the same. However, differences between the groups can be noticed after visualising the data in scatter plots

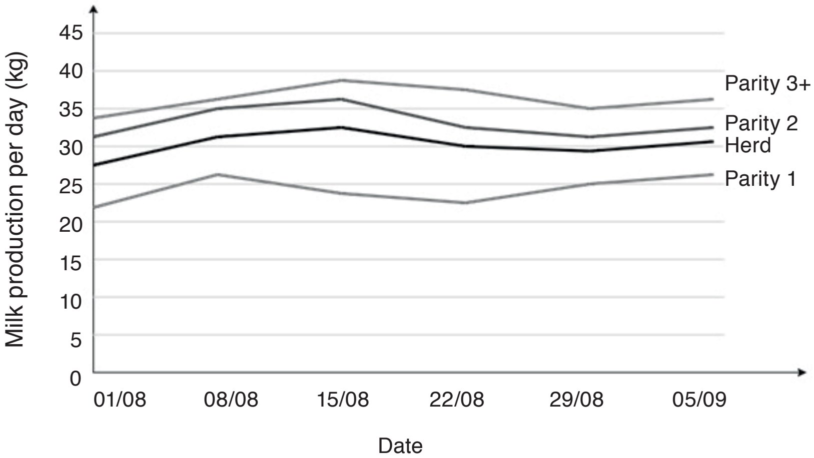

Fig 7: This line graph shows the average daily milk production per parity over a period of five weeks. The cows in third and higher parity are grouped together to avoid an accumulation of lines in the chart. The lines are directly labelled to help practitioners easily understand the graph

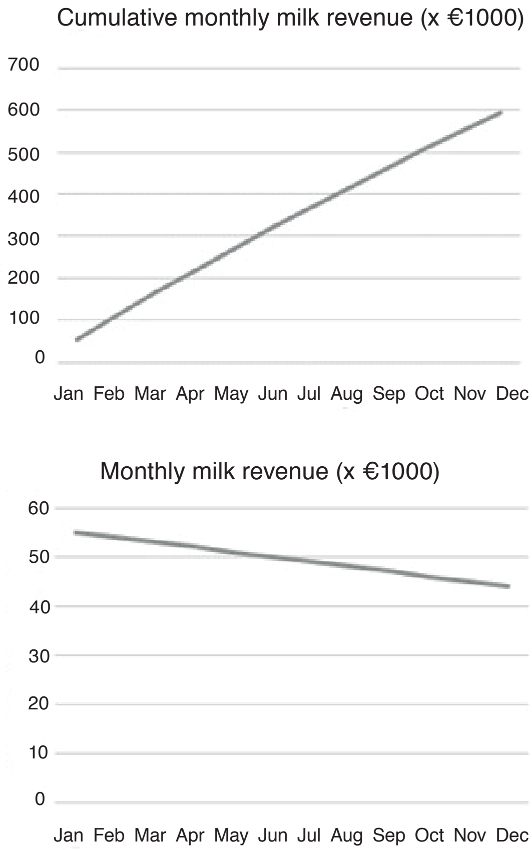

Fig 8: Line graphs showing the monthly milk revenue of a dairy herd. Although the milk revenue per month is decreasing (bottom), the line in the cumulative chart (top) is still going upwards. This may lead to a wrong perception of the actual trend

properties and use of these elements can determine whether a graph will be interpreted correctly or incorrectly (Kelleher and Wagener 2011).

Graph elements

Some graphs contain grid lines as reference points for the existing elements. Deciding whether to use grid lines depends on the data shown on the graph. If the most important feature on the graph is the slope or angle of the graph (eg, a lactation curve), it is not recommended to use grid lines, as they distract the attention of the viewer. A basic rule is to never use grid lines that are darker or more striking than the other elements in the graph (Chambers and others 1983).

A good graph has no unnecessary or distracting elements, it only shows the information that is absolutely necessary. Hence, the use of three-dimensional graphs has no added value above two-dimensional graphs (Barfield and Robless 1989).

Chart types and their use in practice

A variety of graph types exist, each of them being developed with specific data visualisation purposes.

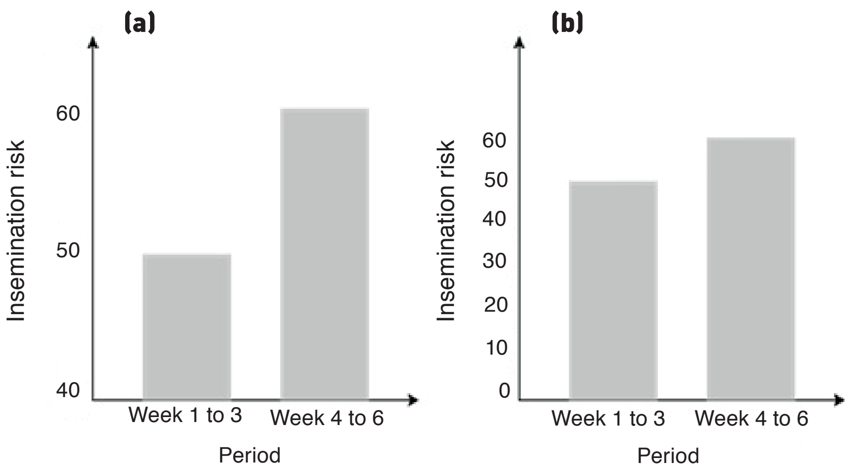

Fig 9: Bar charts demonstrating the effect of deceptive visualisation. Both bar charts show the same data. (a) Due to the fact that the Y-axis does not start at zero, the viewer would think the second bar is double the size of the first bar. (b) In reality, as shown in this correctly designed chart, the difference between the two bars is more nuanced

Line graph

Line graphs are used to show continuous data over a period of time (Kelleher and Wagener 2011). An excellent example of such data is daily milk production. It is advised not to show more than four lines, otherwise it gets difficult for the viewer to interpret the data (Cleveland and McGill 1984). A graph displaying the average milk production for each parity and the average of the total herd (assuming a herd with cows from first to fifth lactation) would contain at least six lines. To avoid this, cows with three or more lactations are commonly grouped into one group, as they have similar milk production capacities (Fig 7), and hence the number of lines in the graph is reduced to four. Lines should be solid and preferably labelled directly and not via a legend (Kelleher and Wagener 2011). One should avoid using cumulative graphs, as these can mask current trends (Fig 8).

Bar chart

Bar charts are used to show and compare discrete data of different groups or time periods (Kosslyn and Chabris 1992). Compared to histograms, bar charts have space between the bars, and this space should be approximately half the thickness of the bar. If a bar's value is zero, then it must be shown in the graph as an empty spot. The axes should be used completely (ie, always start at zero?) (Fig 9), the bars must be distributed equally over the axis and all bars should be the same colour (Rovai and others 2012) (Fig 10).

The order of the bars depends on the KPI shown in the bar charts. If each bar represents a period of time (eg, three weeks in the case of IR), it makes sense to order the bars chronologically. When the bars do not represent a period of time (eg, numbers of cows per country), the bars can be arranged according to their size, either in ascending or descending order. This facilitates the interpretation of the proportions (Kosslyn and Chabris 1992).

Several variations on the classical vertical bar chart exist, such as the horizontal bar chart, the grouped bar chart and the stacked bar chart. In stacked bar charts, each bar contains two or more measurements. It is recommended to limit the amount to two measurements, otherwise it becomes too difficult to compare the different parts. Two KPIs that are often shown together in a stacked bar chart are IR and pregnancy risk (Fetrow and others 1990) (Fig 11).

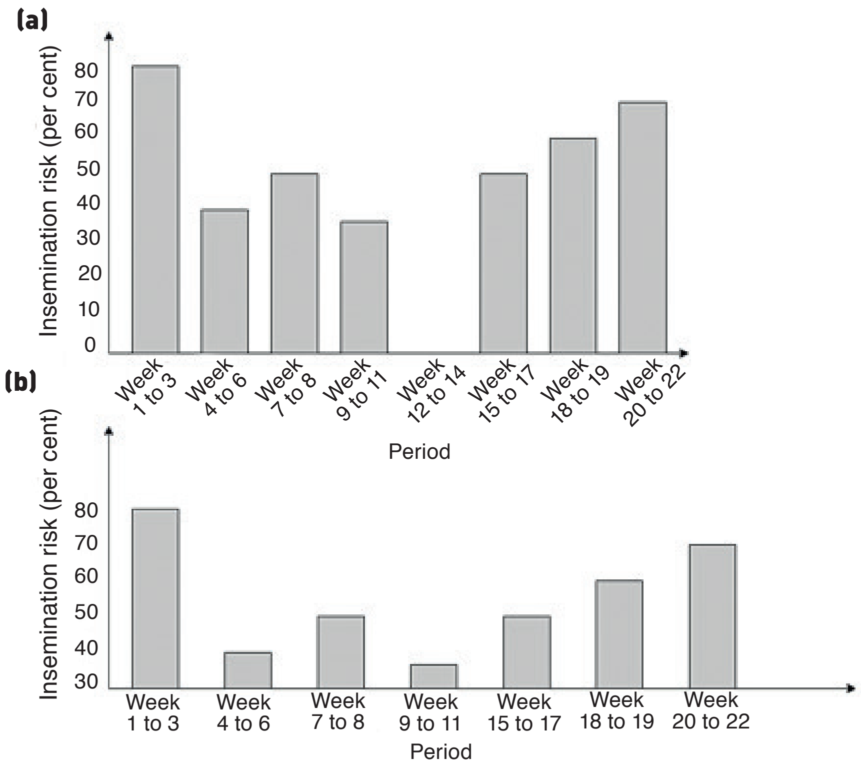

Fig 10: Charts (a) and (b) both present the same data set on the insemination risk in cattle. (a) is designed according to the basic rules when it comes to bar charts. In (b) several mistakes can be seen: the Y axis does not start at zero, which makes it hard to compare the different bars in correct proportions; the space between the bars is too big, the space should be half the thickness of the bars; during week 12 to 14 the insemination risk was zero, but this has been omitted from the chart

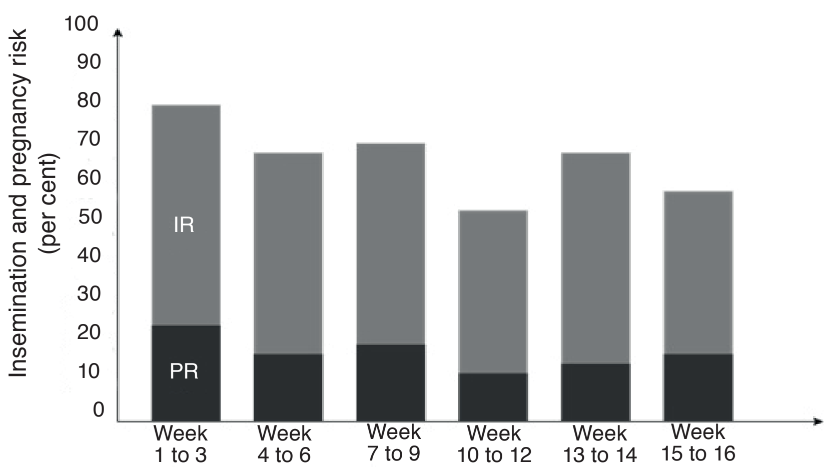

Fig 11: Stacked bar chart showing the insemination risk (IR) and pregnancy risk (PR). The total length of the bar shows the IR, with the dark part only showing the PR

Pie chart

The use of pie charts is strongly discouraged due to their difficulty in interpretation. Quantifying the relative size of the parts is difficult when there are only small differences or when there are too many different parts (Fig 12).

Pie charts are used to visualise proportions of a whole (100 per cent), not to compare different parts. The only case when a pie chart can be used to compare parts is when there are only two parts of a whole (ie, 55 per cent male, 45 per cent female). The use of more than five parts in a pie chart is not recommended as it makes adequate interpretation of the chart difficult (Few 2007) (Fig 12). It is advised to start with the biggest part at '12 o'clock' and proceed clockwise from the biggest to the smallest part. The use of 3D pie charts is not recommended.

Scatter plot

Scatter plots are used to show the correlation between two different parameters (Cleveland 1984). A major advantage of scatter plots is their ability to display individual animal data. Hence, they help to avoid the threat of variation and to detect outliers. An example of a KPI that can be monitored using scatter plots is the amount of days from calving to first insemination (Fig 13).

'Bubbles' or 'circles' are commonly used to display individuals in scatter plots. Adding another parameter to the chart can be done by changing the size of the bubbles. Making the bubbles bigger or smaller can create deceptive visualisation if not executed properly (Pandey and others 2015). It is preferable to size the bubbles according to their area and not to their radius (Fig 14).

Cumulative sum control chart

Cumulative sum control (CUSUM) charts are used for monitoring the mean and variance of a process (Chang and Gan 1995). A CUSUM chart has the advantage of being able to detect smaller and persistent changes in a process. Historic observations are included in CUSUM charts and there is as an upward and downward cumulative sum of observations (de Vries and Conlin 2003).

Insemination success can be evaluated using a CUSUM chart. Every insemination with a known result is plotted chronologically on the chart. If the insemination has been successful, the line goes upwards; if the insemination did not lead to a pregnancy, the line goes downwards. This chart allows vets to have an overview on the conception rate and map this over time. A sudden abnormal downward trend in conception rate, for example, when bovine viral diarrhoea virus infects a herd, can be visualised with this kind of chart.

Radar chart

Radar charts are useful for examining several factors that are related to one item (Kaczynski and others 2008). Radar charts can be used to monitor a KPI over time with its critical limit included, so the reader can quickly detect when the KPI has surpassed the predefined limit (Fig 15).

Conclusion

When presenting KPIs from dairy herds as numbers, issues including variety, bias, lag and momentum arise. Data visualisation can help to handle these issues. A vast variety of graphs exist, and each of them are developed with specific data visualisation purposes. It is important to use them in a correct manner to avoid deceptive visualisation.

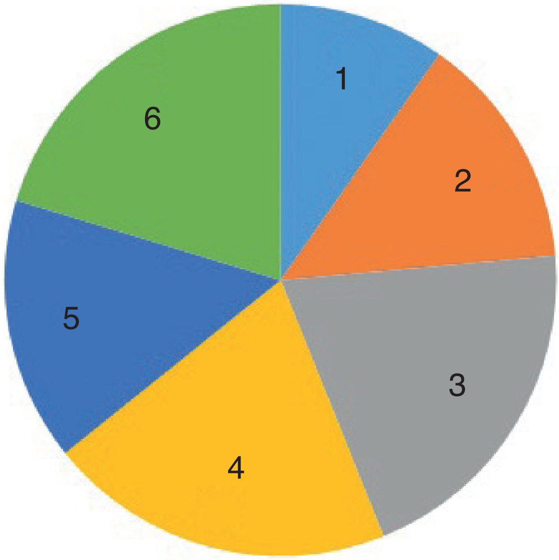

Fig 12: When there are too many parts represented in a pie chart, the proportion between the individual parts becomes unclear, especially when the parts are not ordered in size, as is shown in this pie chart. A pie chart should always be designed with the biggest part at '12 o'clock', with the other parts following clockwise from big to small

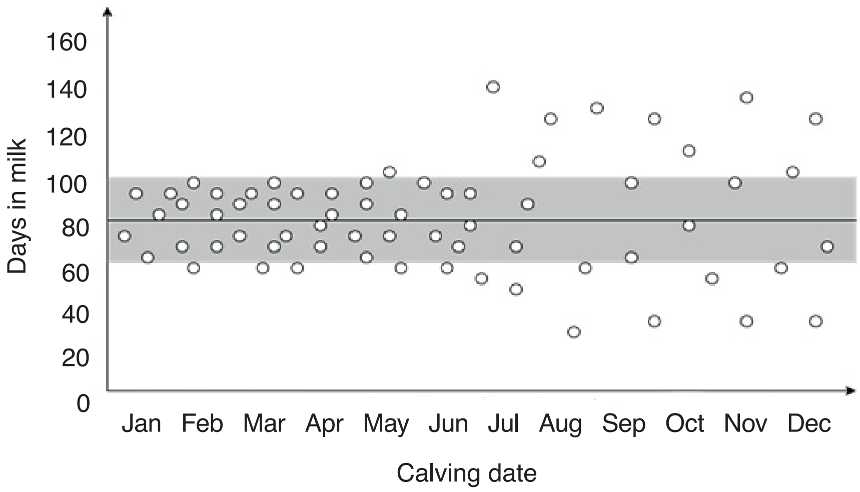

Fig 13: Scatter plot showing the importance of variation monitoring. If we only use the average days in milk (80) as a number without visualising the data, no changes are detected

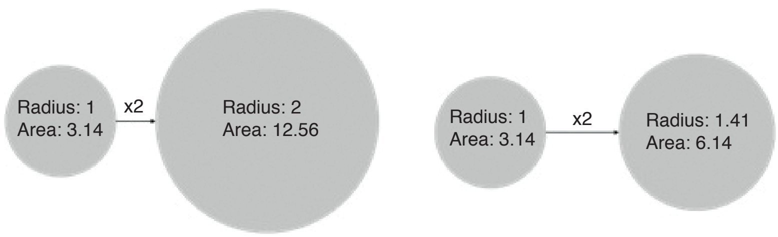

Fig 14: When sizing bubbles in scatter plots, they must be resized according to area and not to radius. This to avoid incorrect proportions

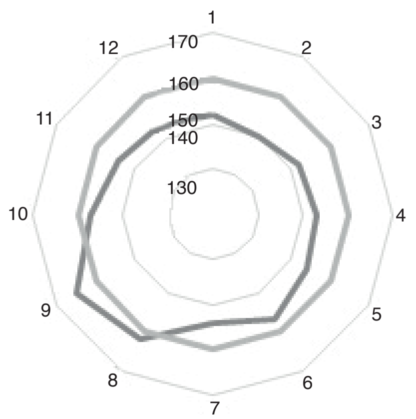

Fig 15: Radar chart showing the average days in milk for each month of the year. These radar charts allow quick detection of a KPI surpassing a certain limit. Here, the limit has been set at 160 days (light grey line). It's easy to see that in month 8 and 9, the average days in milk was too high (dark grey line)

References

ANSCOMBE, F. J. (1973) Graphs in statistical analysis. *The American Statistician* **27,** 17-21

BACH, A. & KERTZ, A. F. (2010) Raising dairy replacements objectively: the value of data-based on-farm decisions. Proceedings of the 19th Annual Tri-State Dairy Nutrition Conference. pp 77-90

BARFIELD, W. & ROBLESS, R. (1989) The effects of two- or three-dimensional graphics on the problem-solving performance of experienced and novice decision makers. *Behaviour and Information Technology* **8,** 369-385

CHAMBERS,J. M., CLEVELAND, W. S., TUKEY, P. A. & KLEINER, B. (1983) Graphical Methods for Data Analysis. Duxbury Press

CHANG, T. C. & GAN, F. F. (1995) A cumulative sum control chart for monitoring process variance. *Journal of Quality Technology* **27,** 109-119

CLEVELAND, W. S. (1984) Graphs in scientific publications. *The American Statistician* **38,** 261-269

CLEVELAND, W. S. & MCGILL, R. (1984) Graphical perception: theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association* **79,** 531-554

DE KRUIF, A. (1978) Factors influencing the fertility of a cattle population. *Journal of the Society for Reproduction and Fertility* **54,** 507-518

DE VRIES, A. & CONLIN, B. J. (2003) Design and performance of statistical process control charts applied to estrous detection efficiency. *Journal of Dairy Science* **86,** 1970-1984

EICKER, S., FETROW, J. & STEWART, S. (2006) Marginal thinking: making money on a dairy farm. *Advances in Dairy Technology* **18,** 137-155

ELLIS, W. (1938) A Source Book of Gestalt Psychology. Routledge and Kegan Paul

FETROW, J., MCCLARY, D., HARMAN, R., BUTCHER, K., WEAVER, L., STUDER E. & OTHERS (1990) Calculating selected reproductive indices: recommendations of the American Association of Bovine Practitioners. *Journal of Dairy Science* **73,** 78-90

FEW, S. (2007) Save the pies for dessert. https://courses.washington.edu/info424/2007/readings/Save%20the%20Pies%20for%20Dessert.pdf. Accessed May 22, 2018

KACZYNSKI, D., WOOD L. & HARDING, A. (2008) Using radar charts with qualitative evaluation: techniques to assess change in blended learning. *Active Learning in Higher Education* **9,** 23-41

KELLEHER, C. & WAGENER, T. (2011) Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling and Software* **26,** 822-827

KOSSLYN, S. M. & CHABRIS, C. F. (1992) Minding information graphics. http://www.chabris.com/Kosslyn1992b.pdf. Accessed May 22, 2018

MANN, C. J. (2003) Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal* **20,** 54-60

MEE, J. & BUCKLEY, F. (2003) Management, nutrition and breeding strategies to improve dairy herd fertility. Proceedings of the National Dairy Conference. pp 134-157

NEBEL, R. L. & DEJARNETTE, J. M. (2007) Reproductive records: monitoring and compliance. Proceedings of the 40th Annual Conference American Association of Bovine Practitioners. pp 50-56

OVERTON, M. W. (2009) Using reproductive records: basics of monitoring. Proceedings of the 46th Florida Dairy Production Conference. pp 33-45

PANDEY, A. V., RALL, K., SATTERTHWAITE, M. L., NOV, O. & BERTINI, E. (2015) How deceptive are deceptive visualizations? An empirical analysis of common distortion techniques. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp 1469-1478

ROVAI, A. P., BAKER, J. D. & PONTON, M. K. (2012) Social Science Research Design and Statistics: A Practitioner's Guide to Research Methods and IBM SPSS Analysis. Watertree Press.

STEINECK, G. & AHLBOM, A. (1992) A definition of bias founded on the concept of the study base. *Epidemiology* **3,** 477-482
VAN DER LEEK, M. L. (2015) Beyond traditional dairy veterinary services:'It's not just about the cows!'. *Journal of the South African Veterinary Association* **86,** 1-10
WONNACOTT, T. H. & WONNACOTT, R. J. (1972) Introductory Statistics. Wiley

Self-assessment quizzes

In Practice has partnered with BMJ OnExamination to host the self-assessment quizzes provided with each clinical article. These can be completed online and found at the end of the online version of each article at www.inpractice.bmj.com

Self assessment: Interpretation and visualisation of data from dairy herds

1. When displaying the average milk production (cows from first to fifth lactation) in a line graph, it is recommended to:
   a. Show the average milk production of all parities (five lines)
   b. Show the average milk production of the herd and all parities (six lines)
   c. To group cows with three or more lactations to limit the amount of lines to four
2. The key performance indicator (KPI) historical calving interval is sensitive for:
   a. The Gestalt effect
   b. The data-ink concept
   c. Lag
3. A cross-sectional calculation measures a KPI at a certain moment in time. True or False?
4. When data (eg, lactation number) is not normally distributed it is recommended to use the mean as the measurement. True or False?

Answers: c, c, true, false