Background: The data displayed pictorially overleaf on page 4.27 were gathered in the 1970s by Dr. J.C. Robinson as one part of a research project, in the Department of Statistics and Actuarial Science at the University of Waterloo, broadly concerned with the adverse health consequences of cigarette smoking. The background to the Question to which these data are relevant is as follows: cigarette smoke is a mixture of thousands of chemicals, but a useful simplification is the binary clas- sification into *tar* and *nicotine*, for which the average levels under defined laboratory conditions have been, for many years in Canada, published on cigarette packages. There is widespread agreement that the numerous adverse health effects of cigarette smoking are associated primarily with chemicals in the tar, whereas the nicotine is the main reason for the 'addictive' nature of cigarette smoking. An important Question was which of the following two strategies is more promising as part of government efforts to reduce the adverse health consequences of smoking in the Canadian population.

* Promote manufacture of cigarettes with *lower* nicotine deliveries, to reduce their 'addictive' nature – the danger is that if individual smokers smoke to attain a given blood nicotine level, this strategy could lead to *in*creased tar intake and an increase in adverse health effects.
* Promote manufacture of cigarettes with *higher* nicotine deliveries, so that smokers could attain a given blood-nicotine level with *lower* intake of tar and, hence, a reduction in adverse heath effects – the danger is that high nicotine cigarettes might be more addictive and hence lead to a higher incidence of smoking and its adverse health effects.

To decide between these two strategies, it was useful to gather data on how individuals smoked with respect to nicotine intake. A way of measuring *mouth*-level nicotine intake was based on the fact that the filter in the tip of a filter cigarette retains a reasonably constant proportion of the nicotine that passes through it as the smoker draws on the cigarette; thus, the nicotine content of the filter provides a measure of how much nicotine the smoker inhaled from that cigarette.

To obtain the data for the display shown overleaf on page HL28.1 (which is for *one* smoker, identified by the letters D.H.T.), Dr. Robinson and his research team identified smokers who would agree to do the following:

- be prepared to smoke, for a number of days, a cigarette brand with a *low* nicotine delivery and, for another period of similar length, a brand with a *high* delivery – these two brands of cigarette were custom made for the investigation;
- record, on a special card provided by the research team, the time of day at which each cigarette was smoked;
- mark each cigarette butt with an appropriate identifier which was also recorded with the time on the card;
- save all their butts each day in a container for the duration of the investigation.

The tasks of the research team were to collect the containers of butts from each participating smoker, dissect out the filter from each butt and then measure its nicotine content. It is clear that obtaining the data for even *one* smoker (as in the display overleaf, which is based on 320 low delivery and 321 high delivery cigarettes smoked by D.H.T.) represents a substantial commitment of expertize and resources. Over a period of about two years, 20 or so smokers participated in this part of the research project.

Each quadrant of the display overleaf shows the distribution of mouth-level nicotine exposure in terms of the length of the time interval between cigarettes; the upper two refer to the cigarette *before* an interval and the lower two to the cigarette *after*. The left-hand half of the display is for the low nicotine delivery brand and the right-hand half for high nicotine delivery. If a smoker is smoking to attain a particular blood nicotine level, we would expect to observe *in*creasing mouth-level exposure with increasing time between cigarettes – the areas with *darker* shading in the lower half of the display should be shifted to the *right* compared with the corresponding part of the upper half; for smoker D.H.T., this does not appear to be the case. Further, it is noteworthy that the low and high nicotine delivery halves of the display appear *similiar* in mouth-level nicotine exposure – that is, for smoker D.H.T., there is little tendency to obtain a *higher* dose of nicotine from the high delivery brand. Thus, these data suggest a strategy of *in*creasing nicotine levels while reducing tar; however, data from other participants in the investigation showed the behaviour of D.H.T. was *not* typical and reducing *both* tar and nicotine levels is actually preferable. [Proper visual gradation among the shadings in the display overleaf requires a printer in good condition; also, gradation quality may degrade under the postscript-to-pdf conversion process and under photocopying of the display.]

NOTE: In the book *The Visual Display of Quantitative Information* (Graphics Press, Cheshire, Connecticut, 1983) Edward R. Tufte gives (p. 51) the following five hallmarks of **graphical excellence**:

- it is the well-designed presentation of interesting information – a matter of *substance*, *statistics* and *design*;
- it consists of complex ideas communicated with *clarity*, *precision* and *efficiency*;
- it gives the viewer the greatest number of ideas in the *shortest time* with the *least ink* in the *smallest space*;
- it is nearly always *multivariate*;
- it requires telling the *truth* about the data.

More generally, these statements should be the goals of *all* data presentation and analysis.

1 Even from the summary given above of Dr. Robinson's investigation, it is clear that a number of *practical* difficulties have significant *statistical* implications; discuss these difficulties and their implications concisely under the headings:

● *study* error; ● *sample* error; ● *measurement* error.