Missing data reduces the statistical power of clinical trials and can lead to elaborate statistical gymnastics to offset the shortcomings of incomplete study databases.1 This article provides a case-in-point and offers suggestions to minimise missing data in future studies.

The United Kingdom Prospective Diabetes Study (UKPDS) was a randomised therapeutic clinical trial in 5,102 subjects with newly diagnosed Type 2 diabetes.2 Subjects attended the clinics every 3-4 months, with comprehensive clinical reviews at entry and thereafter every 3 years. At these triennial clinical reviews, visual acuity (VA) was measured and retinal photographs were taken. A coordinating centre in Oxford, England managed 23 research centres in England, Scotland and Wales. The first subject was randomised in 1977. Median follow-up period was 10 years. Results were first published in the Lancet in 1998.3

We are now using UKPDS study records in a retrospective study of the relationship of lesions of diabetic retinopathy to VA. Our analysis requires at least two consecutive sets of retinal photographs and concurrent VA assessments. With this data, we can measure the affect of changes in diabetic retinopathy lesions on VA.

Retinal photography was introduced into UKPDS in 1982, with funding from the U.S. National Eye Institute (NEI). 22 of the 23 research sites had or obtained suitable equipment. Photographs with 4-field stereoscopic equipment were taken when subjects entered the study and at three-year intervals thereafter for 21 years. Entry photographs could not, of course, be taken of the 540 subjects randomised before 1982, but were taken at the subsequent visits.

Some centres used medical photographers. At others, nurses were trained. Occasionally, a dedicated retinal photographer was available. If the camera was located in the diabetes clinic, photographs could be taken at routine clinical appointments; otherwise, extra appointments were scheduled. As photographs were on transparency (slide) material, the photographers could not immediately see their results and, if needed, take new photographs. Films were processed locally or sent to a single commercial lab, and returned to the coordinating centre.

If photographs did not arrive on schedule, the coordinating center sent reminder letters asking for the photos or an explanation of their absence. Reasons for missing photographs included, for example:

- Subject unwilling to attend
- Subject too unwell to attend extra appointment
- Subject too busy to attend extra appointment ("lambing", "berry harvest")
- Photographer unavailable ("photographer broke ankle", "photographer broke other ankle")
- Camera broken
- Problems with eyes ("dilation", "cataracts")
- Film mailed but not received

At the start of the study, VA was measured using Snellen charts. Based on an unequivocal recommendation by the NEI in the mid 1980’s, the Snellen charts were replaced by LogMAR
charts. The replacement involved more than just swapping one chart for another – the three big new charts for each centre were housed in a box the size of a wardrobe, with internal lighting and concrete bases. 23 of these boxes were built and dispatched. 21 arrived at their destination centre. The location of the other two remains a mystery; those two sites continued with the Snellen charts.

Table 1 presents data on subjects who could have had retinal photographs taken at their three- and six-year visits. From this data, it can be seen that missing photographs were correlated with:

- Younger age
- Heavier weight
- Higher blood pressure
- Smokers when diagnosed with diabetes
- Non-Caucasian ethnic groups

### Table 1. Missing Photographs and Subject Demographics
(Data shown are Mean (s.d.) or n (%))

<table>
<thead>
<tr>
<th></th>
<th>Photos both good quality and graded</th>
<th>Subject did not attend either 3 or 6 year clinic visit</th>
<th>Subject attended 3 and 6 year visits but missing at least 1 photo set</th>
<th>3 and 6 year photos taken but insufficient quality</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>2801</td>
<td>855</td>
<td>681</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>Age at entry to UKPDS (years)</td>
<td>52.5(8.7)</td>
<td>52.0(9.7)</td>
<td>51.5(8.6)</td>
<td>53.7(8.0)</td>
<td>0.0058</td>
</tr>
<tr>
<td>Body Mass Index at entry (kg/m²)</td>
<td>28.9(5.5)</td>
<td>29.6(6.1)</td>
<td>29.2(5.6)</td>
<td>28.2(5.4)</td>
<td>0.0020</td>
</tr>
<tr>
<td>Fasting Plasma Glucose at entry (mmol/L)</td>
<td>12.4(4.3)</td>
<td>12.1(4.1)</td>
<td>12.2(4.2)</td>
<td>12.6(3.9)</td>
<td>0.5952</td>
</tr>
<tr>
<td>Systolic blood pressure at entry (mmHg)</td>
<td>135(21)</td>
<td>137(22)</td>
<td>135(23)</td>
<td>136(20)</td>
<td>0.0475</td>
</tr>
<tr>
<td>Diastolic blood pressure at entry (mmHg)</td>
<td>83(11)</td>
<td>84(12)</td>
<td>83(12)</td>
<td>82(11)</td>
<td>0.0802</td>
</tr>
<tr>
<td>Total cholesterol at entry (mmol/L)</td>
<td>5.6(1.2)</td>
<td>5.6(1.2)</td>
<td>5.6(1.2)</td>
<td>5.5(1.1)</td>
<td>0.5903</td>
</tr>
<tr>
<td>Gender (Men/Women)</td>
<td>1631/1170</td>
<td>536/319</td>
<td>406/275</td>
<td>73/54</td>
<td>0.3340</td>
</tr>
<tr>
<td>Ethnic group (white/Afro Caribbean/Asian Indian/Asian Chinese/other)</td>
<td>2372/166/250/3/10 (64.5%/52.7%/56.8%/33.3%/41.7%)</td>
<td>721/51/75/0/8 (19.6%/16.2%/17.0%/0%/3.3%)</td>
<td>477/91/101/6/6 (13.0%/28.9%/23.0%/66.7%/25.0%)</td>
<td>106/7/14/0/0 (2.9%/2.2%/3.2%/0%/0.0%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Smoking status at entry (never/ex/current)</td>
<td>1003/998/800 (63.4%/65.2%/59.2%)</td>
<td>282/295/323 (16.8%/18.2%/23.0%)</td>
<td>272/228/232 (16.7%/13.1%/15.9%)</td>
<td>48/53/26 (3.0%/3.5%/1.9%)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
The subject population for the VA study is thus significantly different from that of the original UKPDS study, requiring interpretation of the results.

Figure 1 presents availability of VA data. For most centres, 95% or more of the required VA data was in the database. Centre Y did not make measurements reliably on its first subjects. Centre X made no VA measurements in 1988, and performed poorly in the other years. Centre Z became less reliable at the end of the study. Because two consecutive measurements are required for our VA study, a randomly-distributed 15% rate of missing data would mean that only 72% (85% * 85%) of potential analyses could be performed.

For subjects who attended 2 consecutive visits, 71% had both sets of photo data and both sets of VA measurements. 20% had VA measurements for both visits but not photos, 3% had neither photos or VA measurements, and 6% had photos but not VA measurements.

Based on our experience with the study, we can make the following recommendations to ensure completeness of clinical research data:

- Identify risk factors in advance and develop strategies to minimize their effect.
- Closely monitor and analyze completeness of data for factors such as tenure of subject in the study, determine the cause of missing data, and take corrective action immediately.
- Ensure that equipment for measurements is in place at the start of study and is well-maintained during the study, with back-up facilities available.
- Provide competent trainers and verify that equipment operators are competent.
• Ensure that staffing levels are sufficient to cover sickness and holidays.

The use of study data for subsequent retrospective analysis will become more common as new technologies enable researchers to repurpose old data for new uses. For example, it is becoming possible to unfreeze blood samples collected during a study and use them to identify subsets of the study population that disproportionately responded well to the study drug or experienced adverse effects. The study drug can then be marketed to that subset of the patient population (with appropriate screening).

References

Authors
Irene M. Stratton is Senior Medical Statistician in the Diabetes Trials Unit at the Oxford Centre for Diabetes Endocrinology and Metabolism, Churchill Hospital. Contact her at irene.stratton@dtu.ox.ac.uk. Irene Stratton is funded by Diabetes UK (grant 3297).

Stephen J. Aldington, HND DMS FBIPP FinstLM Hon.FOIA, is Director of the Retinopathy Grading Centre, Imperial College London (Hammersmith campus).