# Title

> **pk —** Pharmacokinetic (biopharmaceutical) data

Description     Remarks and examples     References

## Description

The term pk refers to pharmacokinetic data and the Stata commands, all of which begin with the letters pk, designed to do some of the analyses commonly performed in the pharmaceutical industry. The system is intended for the analysis of pharmacokinetic data, although some of the commands are for general use.

The pk commands are

| | | |
|---|---|---|
| pkexamine | [R] **pkexamine** | Calculate pharmacokinetic measures |
| pksumm | [R] **pksumm** | Summarize pharmacokinetic data |
| pkshape | [R] **pkshape** | Reshape (pharmacokinetic) Latin-square data |
| pkcross | [R] **pkcross** | Analyze crossover experiments |
| pkequiv | [R] **pkequiv** | Perform bioequivalence tests |
| pkcollapse | [R] **pkcollapse** | Generate pharmacokinetic measurement dataset |

## Remarks and examples

Several types of clinical trials are commonly performed in the pharmaceutical industry. Examples include combination trials, multicenter trials, equivalence trials, and active control trials. For each type of trial, there is an optimal study design for estimating the effects of interest. Currently, the pk system can be used to analyze equivalence trials, which are usually conducted using a crossover design; however, it is possible to use a parallel design and still draw conclusions about equivalence.

Equivalence trials assess bioequivalence between two drugs. Although proving that two drugs behave the same is impossible, the United States Food and Drug Administration believes that if the absorption properties of two drugs are similar, the two drugs will produce similar effects and have similar safety profiles. Generally, the goal of an equivalence trial is to assess the equivalence of a generic drug to an existing drug. This goal is commonly accomplished by comparing a confidence interval about the difference between a pharmacokinetic measurement of two drugs with a confidence limit constructed from U.S. federal regulations. If the confidence interval is entirely within the confidence limit, the drugs are declared bioequivalent. Another approach to assessing bioequivalence is to use the method of interval hypotheses testing. pkequiv is used to conduct these tests of bioequivalence.

Several pharmacokinetic measures can be used to ascertain how available a drug is for cellular absorption. The most common measure is the area under the time-versus-concentration curve (AUC). Another common measure of drug availability is the maximum concentration ($C_{\max}$) achieved by the drug during the follow-up period. Stata reports these and other less common measures of drug availability, including the time at which the maximum drug concentration was observed and the duration of the period during which the subject was being measured. Stata also reports the elimination rate, that is, the rate at which the drug is metabolized, and the drug's half-life, that is, the time it takes for the drug concentration to fall to one-half of its maximum concentration.

pkexamine computes and reports all the pharmacokinetic measures that Stata produces, including four calculations of the area under the time-versus-concentration curve. The standard area under the curve from 0 to the maximum observed time ($AUC_{0,t_{max}}$) is computed using cubic splines or the trapezoidal rule. Additionally, pkexamine also computes the area under the curve from 0 to infinity by extending the standard time-versus-concentration curve from the maximum observed time by using three different methods. The first method simply extends the standard curve by using a least-squares linear fit through the last few data points. The second method extends the standard curve by fitting a decreasing exponential curve through the last few data points. Finally, the third method extends the curve by fitting a least-squares linear regression line on the log concentration. The mathematical details of these extensions are described in *Methods and formulas* of [R] **pkexamine**.

Data from an equivalence trial may also be analyzed using methods appropriate to the particular study design. When you have a crossover design, pkcross can be used to fit an appropriate ANOVA model. As an aside, a crossover design is simply a restricted Latin square; therefore, pkcross can also be used to analyze any Latin-square design.

There are some practical concerns when dealing with data from equivalence trials. Primarily, the data must be organized in a manner that Stata can use. The pk commands include pkcollapse and pkshape, which are designed to help transform data from a common format to one that is suitable for analysis with Stata.

In the following example, we illustrate several different data formats that are often encountered in pharmaceutical research and describe how these formats can be transformed to formats that can be analyzed with Stata.

▷ Example 1

Assume that we have one subject and are interested in determining the drug profile for that subject. A reasonable experiment would be to give the subject the drug and then measure the concentration of the drug in the subject's blood over a given period. For example, here is a part of a dataset from Chow and Liu (2009, 13):

```
. use http://www.stata-press.com/data/r14/auc
. list, abbrev(14)
```

| | id | time | concentration |
|---|---|---|---|
| 1. | 1 | 0 | 0 |
| 2. | 1 | .5 | 0 |
| 3. | 1 | 1 | 2.8 |
| 4. | 1 | 1.5 | 4.4 |
| 5. | 1 | 2 | 4.4 |
| 6. | 1 | 3 | 4.7 |
| 7. | 1 | 4 | 4.1 |
| 8. | 1 | 6 | 4 |
| 9. | 1 | 8 | 3.6 |
| 10. | 1 | 12 | 3 |
| 11. | 1 | 16 | 2.5 |
| 12. | 1 | 24 | 2 |
| 13. | 1 | 32 | 1.6 |

Examining these data, we notice that the concentration quickly increases, plateaus for a short period, and then slowly decreases over time. pkexamine is used to calculate the pharmacokinetic measures of interest. pkexamine is explained in detail in [R] **pkexamine**. The output is

```
. pkexamine time conc
                                    Maximum concentration =        4.7
                        Time of maximum concentration =              3
                        Time of last observation (Tmax) =           32
                                         Elimination rate =       0.0279
                                                 Half life =      24.8503

    Area under the curve
```

| AUC [0, Tmax] | AUC [0, inf.) Linear of log conc. | AUC [0, inf.) Linear fit | AUC [0, inf.) Exponential fit |
|---|---|---|---|
| 85.24 | 142.603 | 107.759 | 142.603 |

```
    Fit based on last 3 points.
```

Clinical trials, however, require that data be collected on more than one subject. There are several ways to enter raw measured data collected on several subjects. It would be reasonable to enter for each subject the drug concentration value at specific points in time. Such data could be

```
    id   conc1   conc2   conc3   conc4   conc5   conc6   conc7
     1       0       1       4       7       5       3       1
     2       0       2       6       5       4       3       2
     3       0       1       2       3       5       4       1
```

where conc1 is the concentration at the first measured time, conc2 is the concentration at the second measured time, etc. This format requires that each drug concentration measurement be made at the same time on each subject. Another more flexible way to enter the data is to have an observation with three variables for each time measurement on a subject. Each observation would have a subject ID, the time at which the measurement was made, and the corresponding drug concentration at that time. The data would be

```
. use http://www.stata-press.com/data/r14/pkdata

. list id concA time, sepby(id)
```

|      | id | concA    | time |
|------|----|----------|------|
| 1.   | 1  | 0        | 0    |
| 2.   | 1  | 3.073403 | .5   |
| 3.   | 1  | 5.188444 | 1    |
| 4.   | 1  | 5.898577 | 1.5  |
| 5.   | 1  | 5.096378 | 2    |
| 6.   | 1  | 6.094085 | 3    |
| 7.   | 1  | 5.158772 | 4    |
| 8.   | 1  | 5.7065   | 6    |
| 9.   | 1  | 5.272467 | 8    |
| 10.  | 1  | 4.4576   | 12   |
| 11.  | 1  | 5.146423 | 16   |
| 12.  | 1  | 4.947427 | 24   |
| 13.  | 1  | 1.920421 | 32   |
| 14.  | 2  | 0        | 0    |
| 15.  | 2  | 2.48462  | .5   |
| 16.  | 2  | 4.883569 | 1    |
| 17.  | 2  | 7.253442 | 1.5  |
| 18.  | 2  | 5.849345 | 2    |
| 19.  | 2  | 6.761085 | 3    |
| 20.  | 2  | 4.33839  | 4    |
| 21.  | 2  | 5.04199  | 6    |
| 22.  | 2  | 4.25128  | 8    |
| 23.  | 2  | 6.205004 | 12   |
| 24.  | 2  | 5.566165 | 16   |
| 25.  | 2  | 3.689007 | 24   |
| 26.  | 2  | 3.644063 | 32   |
| 27.  | 3  | 0        | 0    |
|      | *(output omitted)* | | |
| 207. | 20 | 4.673281 | 24   |
| 208. | 20 | 3.487347 | 32   |

Stata expects the data to be organized in the second form. If your data are organized as described in the first dataset, you will need to use reshape to change the data to the second form; see [D] **reshape**. Because the data in the second (or long) format contain information for one drug on several subjects, pksumm can be used to produce summary statistics of the pharmacokinetic measurements. The output is

```
. pksumm id time concA
................

Summary statistics for the pharmacokinetic measures
                                          Number of observations =    16
```

| Measure | Mean | Median | Variance | Skewness | Kurtosis | p-value |
|---------|------|--------|----------|----------|----------|---------|
| auc     | 151.63 | 152.18 | 127.58   | -0.34 | 2.07 | 0.55 |
| aucline | 397.09 | 219.83 | 178276.59 | 2.69 | 9.61 | 0.00 |
| aucexp  | 668.60 | 302.96 | 720356.98 | 2.67 | 9.54 | 0.00 |
| auclog  | 665.95 | 298.03 | 752573.34 | 2.71 | 9.70 | 0.00 |
| half    | 90.68  | 29.12  | 17750.70 | 2.36 | 7.92 | 0.00 |
| ke      | 0.02   | 0.02   | 0.00     | 0.88 | 3.87 | 0.08 |
| cmax    | 7.37   | 7.42   | 0.40     | -0.64 | 2.75 | 0.36 |
| tomc    | 3.38   | 3.00   | 7.25     | 2.27 | 7.70 | 0.00 |
| tmax    | 32.00  | 32.00  | 0.00     | .    | .    | .    |

Until now, we have been concerned with the profile of only one drug. We have characterized the profile of that drug by individual subjects by using `pkexamine` and by a group of subjects by using `pksumm`. The goal of an equivalence trial, however, is to compare two drugs, which we will do in the rest of this example.

For equivalence trials, the study design most often used is the crossover design. For a complete discussion of crossover designs, see Ratkowsky, Evans, and Alldredge (1993).

In brief, crossover designs require that each subject be given both treatments at two different times. The order in which the treatments are applied changes between groups. For example, if we had 20 subjects numbered 1–20, the first 10 would receive treatment A during the first period of the study, and then they would be given treatment B. The second 10 subjects would be given treatment B during the first period of the study, and then they would be given treatment A. Each subject in the study will have four variables that describe the observation: a subject identifier, a sequence identifier that indicates the order of treatment, and two outcome variables, one for each treatment. The outcome variables for each subject are the pharmacokinetic measures. The data must be transformed from a series of measurements on individual subjects to data containing the pharmacokinetic measures for each subject. In Stata parlance, this is referred to as a collapse, which can be done with `pkcollapse`; see [R] **pkcollapse**.

Here is a part of our data:

```
. list, sepby(id)
```

| | id | seq | time | concA | concB |
|---|---|---|---|---|---|
| 1. | 1 | 1 | 0 | 0 | 0 |
| 2. | 1 | 1 | .5 | 3.073403 | 3.712592 |
| 3. | 1 | 1 | 1 | 5.188444 | 6.230602 |
| 4. | 1 | 1 | 1.5 | 5.898577 | 7.885944 |
| 5. | 1 | 1 | 2 | 5.096378 | 9.241735 |
| 6. | 1 | 1 | 3 | 6.094085 | 13.10507 |
| 7. | 1 | 1 | 4 | 5.158772 | .169429 |
| 8. | 1 | 1 | 6 | 5.7065 | 8.759894 |
| 9. | 1 | 1 | 8 | 5.272467 | 7.985409 |
| 10. | 1 | 1 | 12 | 4.4576 | 7.740126 |
| 11. | 1 | 1 | 16 | 5.146423 | 7.607208 |
| 12. | 1 | 1 | 24 | 4.947427 | 7.588428 |
| 13. | 1 | 1 | 32 | 1.920421 | 2.791115 |
| | | | | | |
| 14. | 2 | 1 | 0 | 0 | 0 |
| 15. | 2 | 1 | .5 | 2.48462 | .9209593 |
| 16. | 2 | 1 | 1 | 4.883569 | 5.925818 |
| 17. | 2 | 1 | 1.5 | 7.253442 | 8.710549 |
| 18. | 2 | 1 | 2 | 5.849345 | 10.90552 |
| 19. | 2 | 1 | 3 | 6.761085 | 8.429898 |
| 20. | 2 | 1 | 4 | 4.33839 | 5.573152 |
| 21. | 2 | 1 | 6 | 5.04199 | 6.32341 |
| 22. | 2 | 1 | 8 | 4.25128 | .5251224 |
| 23. | 2 | 1 | 12 | 6.205004 | 7.415988 |
| 24. | 2 | 1 | 16 | 5.566165 | 6.323938 |
| 25. | 2 | 1 | 24 | 3.689007 | 1.133553 |
| 26. | 2 | 1 | 32 | 3.644063 | 5.759489 |
| | | | | | |
| 27. | 3 | 1 | 0 | 0 | 0 |
| | | | *(output omitted )* | | |
| 207. | 20 | 2 | 24 | 4.673281 | 6.059818 |
| 208. | 20 | 2 | 32 | 3.487347 | 5.213639 |

This format is similar to the second format described above, except that now we have measurements for two drugs at each time for each subject. We transform these data with pkcollapse:

```
. pkcollapse time concA concB, id(id) keep(seq) stat(auc)
..............................
. list, sep(8) abbrev(10)
```

|     | id | seq | auc_concA | auc_concB |
|-----|-----|-----|-----------|-----------|
| 1.  | 1   | 1   | 150.9643  | 218.5551  |
| 2.  | 2   | 1   | 146.7606  | 133.3201  |
| 3.  | 3   | 1   | 160.6548  | 126.0635  |
| 4.  | 4   | 1   | 157.8622  | 96.17461  |
| 5.  | 5   | 1   | 133.6957  | 188.9038  |
| 6.  | 7   | 1   | 160.639   | 223.6922  |
| 7.  | 8   | 1   | 131.2604  | 104.0139  |
| 8.  | 9   | 1   | 168.5186  | 237.8962  |
| 9.  | 10  | 2   | 137.0627  | 139.7382  |
| 10. | 12  | 2   | 153.4038  | 202.3942  |
| 11. | 13  | 2   | 163.4593  | 136.7848  |
| 12. | 14  | 2   | 146.0462  | 104.5191  |
| 13. | 15  | 2   | 158.1457  | 165.8654  |
| 14. | 18  | 2   | 147.1977  | 139.235   |
| 15. | 19  | 2   | 164.9988  | 166.2391  |
| 16. | 20  | 2   | 145.3823  | 158.5146  |

For this example, we chose to use the AUC for two drugs as our pharmacokinetic measure. We could have used any of the measures computed by pkexamine. In addition to the AUCs, the dataset also contains a sequence variable for each subject indicating when each treatment was administered.

The data produced by pkcollapse are in what Stata calls wide format; that is, there is one observation per subject containing two or more outcomes. To use pkcross and pkequiv, we need to transform these data to long format. This goal can be accomplished using pkshape; see [R] **pkshape**.

Consider the first subject in the dataset. This subject is in sequence one, which means that treatment A was applied during the first period of the study and treatment B was applied in the second period of the study. We need to split the first observation into two observations so that the outcome measure is only in one variable. Also we need two new variables, one indicating the treatment the subject received and another recording the period of the study when the subject received that treatment. We might expect the expansion of the first subject to be

| id | sequence | auc | treat | period |
|----|----------|----------|-------|--------|
| 1  | 1        | 150.9643 | A     | 1      |
| 1  | 1        | 218.5551 | B     | 2      |

We see that subject number 1 was in sequence 1, had an AUC of 150.9643 when treatment A was applied in the first period of the study, and had an AUC of 218.5551 when treatment B was applied.

Similarly, the expansion of subject 10 (the first subject in sequence 2) would be

| id | sequence | auc | treat | period |
|----|----------|----------|-------|--------|
| 10 | 2        | 137.0627 | B     | 1      |
| 10 | 2        | 139.7382 | A     | 2      |

Here treatment B was applied to the subject during the first period of the study, and treatment A was applied to the subject during the second period of the study.

An additional complication is common in crossover study designs. The treatment applied in the first period of the study might still have some effect on the outcome in the second period. In this example,

each subject was given one treatment followed by another treatment. To get accurate estimates of treatment effects, it is necessary to account for the effect that the first treatment has in the second period of the study. This is called the carryover effect. We must, therefore, have a variable that indicates which treatment was applied in the first treatment period. pkshape creates a variable that indicates the carryover effect. For treatments applied during the first treatment period, there will never be a carryover effect. Thus the expanded data created by pkshape for subject 1 will be

```
id   sequence   outcome   treat   period   carry
 1          1   150.9643       A        1       0
 1          1   218.5551       B        2       A
```

and the data for subject 10 will be

```
id   sequence   outcome   treat   period   carry
10          2   137.0627       B        1       0
10          2   139.7382       A        2       B
```

We pkshape the data:

```
. pkshape id seq auc*, order(ab ba)
. sort id sequence period
. list, sep(16)
```

|     | id | sequence | outcome  | treat | carry | period |
|-----|----|----------|----------|-------|-------|--------|
| 1.  | 1  | 1        | 150.9643 | 1     | 0     | 1      |
| 2.  | 1  | 1        | 218.5551 | 2     | 1     | 2      |
| 3.  | 2  | 1        | 146.7606 | 1     | 0     | 1      |
| 4.  | 2  | 1        | 133.3201 | 2     | 1     | 2      |
| 5.  | 3  | 1        | 160.6548 | 1     | 0     | 1      |
| 6.  | 3  | 1        | 126.0635 | 2     | 1     | 2      |
| 7.  | 4  | 1        | 157.8622 | 1     | 0     | 1      |
| 8.  | 4  | 1        | 96.17461 | 2     | 1     | 2      |
| 9.  | 5  | 1        | 133.6957 | 1     | 0     | 1      |
| 10. | 5  | 1        | 188.9038 | 2     | 1     | 2      |
| 11. | 7  | 1        | 160.639  | 1     | 0     | 1      |
| 12. | 7  | 1        | 223.6922 | 2     | 1     | 2      |
| 13. | 8  | 1        | 131.2604 | 1     | 0     | 1      |
| 14. | 8  | 1        | 104.0139 | 2     | 1     | 2      |
| 15. | 9  | 1        | 168.5186 | 1     | 0     | 1      |
| 16. | 9  | 1        | 237.8962 | 2     | 1     | 2      |
| 17. | 10 | 2        | 137.0627 | 2     | 0     | 1      |
| 18. | 10 | 2        | 139.7382 | 1     | 2     | 2      |
| 19. | 12 | 2        | 153.4038 | 2     | 0     | 1      |
| 20. | 12 | 2        | 202.3942 | 1     | 2     | 2      |
| 21. | 13 | 2        | 163.4593 | 2     | 0     | 1      |
| 22. | 13 | 2        | 136.7848 | 1     | 2     | 2      |
| 23. | 14 | 2        | 146.0462 | 2     | 0     | 1      |
| 24. | 14 | 2        | 104.5191 | 1     | 2     | 2      |
| 25. | 15 | 2        | 158.1457 | 2     | 0     | 1      |
| 26. | 15 | 2        | 165.8654 | 1     | 2     | 2      |
| 27. | 18 | 2        | 147.1977 | 2     | 0     | 1      |
| 28. | 18 | 2        | 139.235  | 1     | 2     | 2      |
| 29. | 19 | 2        | 164.9988 | 2     | 0     | 1      |
| 30. | 19 | 2        | 166.2391 | 1     | 2     | 2      |
| 31. | 20 | 2        | 145.3823 | 2     | 0     | 1      |
| 32. | 20 | 2        | 158.5146 | 1     | 2     | 2      |

As an aside, crossover designs do not require that each subject receive each treatment, but if they do, the crossover design is referred to as a complete crossover design.

The last dataset is organized in a manner that can be analyzed with Stata. To fit an ANOVA model to these data, we can use anova or pkcross. To conduct equivalence tests, we can use pkequiv. This example is further analyzed in [R] **pkcross** and [R] **pkequiv**.

◁

# References

Chow, S.-C., and J.-P. Liu. 2009. *Design and Analysis of Bioavailability and Bioequivalence Studies.* 3rd ed. Boca Raton, FL: Chapman & Hall/CRC.

Ratkowsky, D. A., M. A. Evans, and J. R. Alldredge. 1993. *Cross-over Experiments: Design, Analysis, and Application.* New York: Dekker.