Protocols for Data Management

Altmann Lab – Princeton Work

Updated November 2015

CONTENTS

GENERAL INTRODUCTION

This document contains all necessary procedures for data management at Princeton University related to the Amboseli Baboon Project. This includes a description of procedures for handling incoming data from the field as well detailed guidelines on how to enter and update information into the master database.

Your work will go smoothly if you keep in mind a few basic rules:

1. *Always* back up your work.
2. Follow the system backup schedule.
3. Keep up with the data entry and proofing.
4. Enter and save Amboseli data in the appropriate subdirectories of C:\Database Files\Altmann.
5. Keep track of all proofing using the "Proofed" comment boxes in the Excel files.
6. Keep track of all updating by sending an announcement on the babase mailing list and updating the Status of Babase Datasets table on the wiki homepage.
7. Keep your own work in My Documents.
8. Document any new protocols or protocol changes.
9. Don't panic!

Most of the data management work should follow these steps:

1. Sorting/filing of incoming data from Amboseli
2. Data entry
3. Data proofing and correcting
4. Data validating and updating
5. Data archiving

The following sections of the protocol will explain in detail how these steps are followed with the different datasets. Keep in mind that the goal of this manual is *not* to provide information on the Babase data management system itself or to give detailed descriptions of the types of data and data collection methods in the field. Further information on these topics can be found in the *Amboseli Baboon Project: Data Management System* and the *Guidebook for the Long-term Monitoring of Amboseli Baboons and their Habitat*, respectively.

FILING AND MANAGEMENT OF AMBOSELI DATA

Until 2014 a package containing photocopies of the original data collected in Amboseli arrived every month in the mail. These photocopies were copied. The second copy was filed and stored in Jeanne's lab at Princeton, while the copies from Amboseli were sent to Duke. Beginning with the Dec 2013 data, data are scanned in Nairobi and emailed to both database managers and the baboon leaders. These files are saved at Princeton on the database manager's computer in the electronic data file for the appropriate month (see the section of archiving of electronic data). Those items needed here are printed and filed in

the normal manner. The original data sheets remain in Amboseli until someone returning from the field personally brings them back for filing and storage in Susan's lab at Duke.

Filing of Photocopies
Each package of data should include a "Data Sent List." Inventory the data by sorting through the package contents and checking off items on the "Data Sent List" as you find them. Be sure to make note of any unreadable photocopies and/or any missing data and immediately request these sheets be resent from Amboseli.

When you are finished doing the inventory of the data, place the "Data Sent List" in the appropriate Tupperware bin beside the door to the fecal prep lab and file the package contents as follows:

> *Monitoring Data* – There is a monitoring notebook for each study group for each calendar year. The following types of data are filed in the monitoring notebooks: group census, demography notes, sex-skin scoring, reproductive notes, subgroups, mounts/consorts, other groups, predation/human disturbance, grooming, decided agonisms, undecided agonisms, multiparty interactions, wounds and pathologies, and other notes.

> *Meteorological Data* – Hand-recorded data on daily min/max ambient temperature and rain gauge readings at the field camp are sent each month. These are filed chronologically in the "Meteorological Data" notebook. *Note that meteorological data are also recorded by a WeatherHawk weather station set up at camp – we receive this electronic data as CSV files on a weekly basis as an email attachment from Amboseli.*

> *Charcoal Fridge Temperature* – Min/max data for the temperature of the charcoal fridge is sent each month. These data are filed chronologically in the notebook called "Charcoal Fridge Temperatures and Half-hourly Temperatures."

> *Neonatal Assessments* – There are two Neonatal Assessments notebooks currently used for the offspring of living females: one for groups descended from Alto's group (Viola's, Hokey's, Snap's, and Acacia's – including the former members of Laza's group though they are descended from Hook's females) and one for groups descended from Hook's group (Narasha's, Mica's, and Kelly's). Note that, following a series of group fissions, Kelly's group ceased to be monitored in July 2012 and Snap's and Mica's groups ceased to be monitored in Dec 2012. The notebooks are organized by group, mother's name, and finally offspring birth order. As you file a new neonatal assessment, mark it with a flag labeled with the infant's sname – this will help Jeanne quickly locate the new births when later assigning birth dates. Color code the tags by update period.

> In addition to these binders, there is one notebook for all neonatal assessments of infants whose mothers are from the Alto's or Hook's lines but are now dead. When a female dies, be sure to move the neonatal assessments of all her offspring

to this notebook - the best time to do this is during the demography update. Lastly, there are two separate neonatal assessment notebooks for the Lodge groups, one with infants of mothers still alive at the end of monitoring of those groups and one for those whose mothers were already dead at that time.

These data were made available in Babase_pending beginning in 2013. The table neonatals is updated during the regular quarterly Babase update.

Canine Condition and Scrotal Development – These two data types are filed together in a single notebook organized by data type, group, and year. There is much repetition in what is sent from Amboseli with regard to these datasets – be certain to keep up-to-date versions only and discard redundant older versions of the data.

Hybridity Scoring – Hybridity scoring sheets are organized chronologically by group in the Hybridity Scores notebook. Hybridity was originally handled at Duke but was passed to Princeton ca. 2011. Raw data from 2000 through 2013 were added to Babase_pending in Feb 2014. Henceforth the table hybridity_rawmorpho will be updated during the quarterly Babase update. This table joins the hybridmorpho (derived from the raw morphology scores determined in the field) and hybridgene (derived from genetic data) tables already available in babase_pending.

Male Age Estimates – Male age estimate sheets are organized in reverse chronological order (regardless of baboon's name and natal group) in a notebook called Male Age Estimates. This binder included initial age estimate sheets for newly immigrated males, as well as annual age reassessments of all adult males. The tables are updated in Excel monthly as new age estimates arrive. Currently these data are not available anywhere in Babase.

Other Groups – When possible, census data are collected for several other baboon groups in the Amboseli region including groups that have never been study groups and groups that have been dropped as the population and number of groups have increased. These include Snap's, Mica's, Kelly's, Ositeti (Jill's and Ceejay's), Olkenya, the Sinya groups, Joy's, Nzige's, Proton's, and Stud's groups. These data are organized by group and date in the "Other Groups" notebooks.

Collar Checks – Sheets recording when baboon collars were listened for and which ones were heard are filed in the collar checks notebook. Beginning at the end of June 2013 new collar check sheets on baboons in other groups (including recently dropped study groups) were added to help keep better track of these baboons. These checks are performed by drivers during the other groups census day and during the last days. The sheets include space for recording notes on where the baboons were seen and demography notes may be derived from these sheets. They are filed in a separate section of the collar checks notebook.

Morphometrics/Frozen Fecals – These data are filed in chronological order in the "Darting" notebook.

Tree Grove Monitoring – These data are filed in chronological order in the "Tree Grove Monitoring" notebook.

Bills and Receipts – Give any original copies of bills and receipts to Jeanne's assistant for accounting.

Phone Log – Give any copies of phone logs to Jeanne's assistant for filing. The Team stopped sending this in the middle of 2013.

Daily Activity Calendars – Alphabetize monthly calendars by employee's first name and staple together in one monthly packet. We have also been receiving a summary activity calendar listing which groups were observed on each monitoring day. This summary calendar is stapled as the first calendar in the monthly packet. File in the appropriate Tupperware bin beside the door to the fecal prep lab. In 2013 RTI sheets were also received and were filed with the activity calendars. These are now filed electronically and not printed.

Amboseli Log Sheets – Beginning in 2006, the field Team sends copies of various log sheets. Some of these log sheets serve as useful documentation of which file(s) were sent (in the case of datasets that are received electronically from the field), some are useful for keeping track of sampling schemes (in the case of focal animal logs), and some are simply documentation that data are being backed-up on the camp computer (in the case of the Dell A log). These are now filed electronically and not printed.

Disks from Amboseli – In the past, disks arriving from Amboseli contained a variety of data files and electronic datasets. In more recent times, much of this information is sent directly to Princeton and Duke via email from the Team. As a result, disks arrive far less frequently than they did in the past. One notable exception is the arrival of occasional digital photo CDs. Refer to the "digital photos" section of this protocol guide for notes on managing and cataloging images. The disk itself is stored beside the fax machine in the CD rack in Jeanne's lab. Note that disks with various contents are stored in different colored jewel cases. No disks have been sent in many years.

Log of Vehicle Trips/Repairs - In 2008, the Team began sending a log of the vehicle trips and repairs in an attempt to better anticipate servicing needs and keep more vehicles on the road. A copy of these should be sent to Duke and a copy should be filed in the Vehicle Reports binder. These are now filed electronically and not printed.

Review of Photocopies
If time allows, it is helpful to take a few minutes to read through the photocopied data right away to get a feel for what is going on. This will alert you to any major group changes and you may notice some mistakes that can be corrected quickly.

Mailing Data Copies to Duke
Note: This was the procedure for handling photocopied data – it was used through 2013 and should be used again should a reversion to paper data be required in future.

Photocopies are sent monthly to Princeton. The originals arrive at Duke only after Jeanne or Susan makes a trip to Kenya and hand-carry them back again. Princeton is therefore responsible for getting recent copies of the data to the database manager at Duke. In the past we sent ad lib data (agonisms, mounts and consorts, and grooming files) as well as study-group censuses, and demography notes to Susan's lab for data entry, proofing, and updating. We also used to send all log sheets relevant to data management at Duke (i.e. information pertinent to GPS and Psion datasets) as well as activity calendars and vehicle logs. However, in March 2008 the xeroxing in Nairobi had gotten so erratic that the Team began making the copies on the 3-in-1 machine at camp. This machine has a type of non-permanent ink so the Kenyan copies we now receive from the field Team each month are not suitable for archiving. Consequently, the entire data packet must be copied on a Princeton machine and **those copies kept at Princeton**. The data sheets from Kenya now get sent to Duke via regular U.S. mail. Originals are still carried to and stored at Duke.

Emailed Files
In addition to the photocopied data arriving in the mail, the Amboseli field team also emails electronic data files on a weekly basis. These weekly files contain three main types of data: meteorological data collected by the WeatherHawk station (.dat, formerly.csv, using the old WeatherHawk software), point sample data (.pts), and GPS readings (.mps). This email is sent to both Princeton and Duke: Princeton works with the meteorological data, while Duke works with the point sample data and the GPS readings. When a new email arrives, unzip the attached files and save only the meteorological data to the appropriate month/year folder in C:\Database Files\ALTMANN\Data from Amboseli\ (refer to the section in this protocol guide on meteorological data for more information on how to work with these files). Duke is responsible for downloading and saving the point sample and GPS data.

The Team also emails monthly summary files. These typically include Excel spreadsheets for salary calculation and cash accounting and a Word document for the monthly report. Forward the salary calculation and cash account files to Jeanne's assistant for accounting and save the original files in the appropriate subfolder in C:\Database Files\ALTMANN\Data from Amboseli\. Print out a hard copy of the monthly report and file it in the Tupperware bin beside the fecal prep lab. Save the original file in the appropriate subfolder in C:\Database Files\ALTMANN\Data from Amboseli\. Additional

files such as the agonism matrices and field-entered interaction data may be sent separately.

FILE NAMING CONVENTIONS FOR THIS DOCUMENT

Whenever a table from Babase 2.0 is named in this document, it will appear in all caps, such as BIOGRAPH.

Whenever a Babase 2.0 data maintenance program is named in this document, it will appear in italics, such as *Ranker*.

Specific file names will appear in **bold**.

Column headers specific to the demography update sheets (the paper sheets we use to summarize data for the update) will be named with an underline (e.g. infant sname)

Whenever this document gives directions for doing certain tasks in software programs, the software menu options, buttons, etc for that program will be shown in red text with a greater than sign (>) to show the submenu sequence choices you must make.

Any important information/notes will be written using the **Impact Font.**

NAMING CONVENTIONS FOR DEMOGRAPHY UPDATE FILES

To upload data into the Babase 2.0 database (also referred to as 'Babase' in this document), we generally create and populate Excel files, export them to tab delimited text format and upload. Most demography update files are currently stored on the Princeton database manager's computer in:
C:\Database Files\ALTMANN\Babase\TEMP\
within subfolders for **Repro** and **Census**.

There are also many electronic matrices exported from the program *Ranker;* these we store on the Princeton database manager's computer in: C:\Database Files\ALTMANN\Babase\TEMP\Ranks.

The update files follow the same general naming conventions whether files were created for import into Babase 1.0 or Babase 2.0. When Babase 1.0 was in use, all files were created as .dbf tables for upload to the FoxPro program. With every .dbf file upload, the program generated a .txt output file to document the update. Theses .txt files are stored with the .dbf files in the folders named above.

In Babase 2.0, the upload files are prepared in .xls format for ease of input and proofing, but then exported to a .txt format for ease of upload. There are no .txt output files from the uploads that occur with Babase 2.0; however, an upload log is kept by the database manager using Word to document the update process and errors encountered.

This is the generalized format for naming the various upload files you may see stored in the TEMP folder:

G T M M Y Y a/b .dbf
G T M M Y Y a/b/c/d .xls
G T M M Y Y a/b/c/d .txt

G = Group and refers to a one letter abbreviation of the population of study animals who were observed to produce the data in the dataset. The groups most often referenced in file naming are **L**inda's, **N**yayo's, **O**mo's, **V**iola's, **W**eaver's, **H**okey's, Na**r**asha's, **D**ibble's, and Vo**g**ue's. Other study groups include **A**lto's, **D**otty's, **J**oy's, **L**odge, N**z**ige's (for 1990s data), **K**elly's, Laza's, **S**nap's, **and M**ica's (for 2011-2012). Censuses of nonstudy groups are coded with a **G**.

T or **TT** = Type of data. This field usually consists of one or two digits. The types of data that were previously entered regularly at Princeton were **A**gonism, **Bir**ths, Ce**n**sus, **D**emography, **G**rooming, Mounts/**C**onsorts/Ejaculations, **P**regnancies, and **R**eproduction. (In the case of ranks, in this field **t** = temporary and **f** = final to designate the temp or finalized matrices). Duke now enters the Agonism, Grooming, and Mount/Consort data so these headings will only be seen in older data files. Female **cyc**ling, **m**aturity, and **m**ale **dis**persals are also entered at Princeton.

MM = Two digits denoting the month. Always use a leading zero for months with MM less than 10. Use **01** for January, **02** for February, **03** for March, etc.

YY = Two digits denoting the year. Always use leading zeros for years with YY less than 10. Use **99** for 1999, **00** for 2000, **01** for 2001, etc.

a or **b** = This notation is used in different ways depending on the file type. When used in conjunction with ad lib datasets, where the proofing system requires entry of two sets of data, the **a** is used for the first entered set and the **b** is used to denote the second entered set. Again, these data are now entered at Duke so we will only find such files in records for older data uploads.
In the case of files directly associated with demography updates, **a** and **b** were used to differentiate files relevant to the two demography updates of the calendar year (i.e. Jan-June 2007 = "07a" and July-Dec = "07b"). This was the case through 2011. Halfway through 2011, we switched to quarterly updates. In the second half of 2011, however, these were not yet true quarterly updates because reproductive data (pregs and cycling) were not uploaded and the two 11b updates were referred to as 11b(1) and 11b(2).

a, b, c, and **d** = Beginning in 2012 we employed true quarterly updates where all data were updated on a quarterly basis. Thus a = quarter 1 (Jan – Mar), b = quarter 2 (Apr – Jun), c = quarter 3 (Jul – Sep), and d = quarter 4 (Oct – Dec).

A few quick notes on why there may be extra or different files in the demography update folders from update to update:

1) For Babase 2.0 input files, there should be an exact match between the data in files created in .xls and those exported from .xls to .txt for uploading. However, always check the .txt file as the definitive source of what actually went into the database because only .txt files can be uploaded.

2) The FoxPro database protocol was retired just after the Princeton 2006b update, but prior to the Duke 2006b update. Duke's first update in Babase 2.0 was done for 2006b data (psion, ad libs, agonisms). Princeton's first demography update in Babase 2.0 was done for 2007a. Thus, there should be .dbf files for Princeton through all of 2006. A new file and folder format appears for the 2007a upload.

3) The 07a update revealed an *Upload* program bug: no "Demography" notes or "Other Group" notes would upload to the DEMOG_CENSUS view whenever there is a mismatch between the value in the "grp" column and the value in the "reference" column (records with a match seem to upload). The workaround for this bug required the creation of extra .txt files to update the CENSUS and DEMOG tables separately, since it cannot be done through the view. The workaround is documented in a **ReadMe** in: C:\Database Files\ALTMANN\Babase\TEMP\CENSUS

This error in the *Upload* program code has since been identified and corrected. Workarounds should not be necessary for subsequent demography updates and the numerous .xls and .txt files seen in the folders for **Demog07a** and **Other07a** will not occur in **Demog07b**, **Other07b**, or later update folders.

4) Files for census backfill for Hook's and Alto's (to replace the members-style fake census data for a portion of the project's earlier data) can be found in C:\Database Files\ALTMANN\Babase\TEMP\CENSUS\census backfill.

5) Files for demography note backfill (which will be ongoing for quite some time) can be found in C:\Database Files\ALTMANN\Babase\TEMP\CENSUS\demog note backfill. Files regarding review of death and dispersal information related to backfill periods can also be found in this folder, along with a demog note backfill log. Demog note backfill and related deaths and dispersals are organized by group. Check the Status of Babase Datasets table on the wiki's homepage for the current status of the demography backfill effort.

6) Files for changes to census data related to male identity consolidations (two males that turned out to be one male), the Edgy/Eclipse switcheroo, etc. can be found in C:\Database Files\ALTMANN\Babase\TEMP\CENSUS\Identity Alterations.

7) Files containing information on assignment of death and dispersal dates and dcauses, and death and dispersal confidence levels (beginning with the 09a update with confidence

levels added sometime in 2011), as well as files concerning initial censoring of individuals no longer within the study population (Nov 2009) and backfill confirmed dispconfidences (May 2011) can be found in C:\Database Files\ALTMANN\Babase\TEMP\CENSUS\deaths and dispersals.

8) Information organized for the April 2010 review of demography note protocols can be found here: C:\Database Files\ALTMANN\Babase\TEMP\CENSUS\demog note backfill\Demog note revamp. This marked a shift away from recording of nondemographic behavior in the demography notes. Behavior items and other miscellaneous, nondemographic items are now indexed and the index for each update can be found in the final tab of the demography notes file for each update period.

DEMOGRAPHY UPDATES

A "demography update" refers to consolidating the following hard data into computer files that are subsequently uploaded into the Babase 2.0 (online) database tables and views. You will generally need at least one month's data beyond the target time period to complete the demography update. The update can be done without this information but the final month's pregnancies will generally not be available and cycling data will have a wider margin for error at the end of the update period, as repstats and cycstats will be inaccurate for longer periods of time for individuals that have changed state toward the end of the update period.

Table describing where and how to upload each type of data that is updated at Princeton.

Hard data	Input program/ method	Upload location	Table or view	Main tables affected	Views affected
Births	upload program	BIOGRPAH	Table	BIOGRAPH	BIOGRAPH_SORTED MATERNITIES
Aborts	upload program	BIOGRAPH	Table	BIOGRAPH	BIOGRAPH_SORTED MATERNITIES
Immigrant M	upload program	BIOGRAPH	Table	BIOGRAPH	BIOGRAPH_SORTED
Census	upcen program	CENSUS	Table	CENSUS BIOGRAPH MEMBERS	CENSUS_DEMOG (sorted) DEMOG_CENSUS (sorted)
F repro cycles	upload program	MTD_CYCLES	View	CYCPOINTS CYCLES PREGS MMINTERVALS MDINTERVALS CYCSTATS* REPSTATS*	MTD_CYCLES MATERNITIES CYCLES_SEXSKINS (sorted)
Pregnancies	upload program	PREGS	Table	PREGS	MATERNITIES
Maturedates F	upload	MATUREDATES	Table	MATUREDATES	

	program				
Maturedates M	upload program	MATUREDATES	Table	MATUREDATES	
Demog notes	upload program	DEMOG_CENSUS	View	CENSUS DEMOG	CENSUS_DEMOG (sorted) DEMOG_CENSUS (sorted)
Other groups	upload program	DEMOG_CENSUS	View	CENSUS DEMOG	CENSUS_DEMOG (sorted) DEMOG_CENSUS (sorted)
Deaths	SQL Update	BIOGRAPH	Table	BIOGRAPH	BIOGRAPH_SORTED
Dispersal M	upload program	DISPERSEDATES	Table	DISPERSEDATES	
Sex skin size	upload program	SEXSKINS_CYCLES	View	SEXSKINS	SEXSKINS_CYLCES SEXSKINS_CYCLES (sorted)
PCS color	upload program	PCSKINS	Table	PCSKINS	SPCSKINS (sorted)
Manual weather data	upload program	MIN_MAXS	View	RAINGUAGES TEMPMAXS TEMPMINS WREADINGS	MIN_MAXS MIN_MAXS (sorted)
WeatherHawk weather data	upload program	Currently in Babase_pending			
Wounds and pathologies	upload program	Currently in Babase_pending			
Neonatals	upload program	Currently in Babase_pending			

* table change depends upon running the "rebuild" program for these tables.

There should rarely be direct manual input into the tables of online database tables during these updates. **The update files should always be uploaded to Babase 2.0 through the data maintenance programs.** These programs are referred to as "**Babase's bespoke programs"** in the papio index or the wiki). However, if small manual changes are necessary to correct errors caught outside the parameters of the demography updates, you can make these without worry (or use of the bespoke programs). Babase 2.0 should catch most errors and not allow the row alteration if it finds conflicting data in another table. The table below summarizes the input procedure for various types of data.

In the past, doing a demography update for a six-month time period has worked well by
(1) providing enough consecutive months to limit "edge effects" and
(2) keeping the amount of work involved manageable.

In recent years the data have arrived much more quickly from the field and have been entered upon arrival. This has made the switch to quarterly updates quite viable.

There are 8 major steps in any demography update.
Step 1 - Manual data review and consolidation
  1a: Jeanne's review of female sex skin data
  1b: Copy datasheet info to handwritten demography update sheets
  1c: Email Team about uncertain pregnancies, new individual names and sexes
  1d: Evaluate disappearances to assign death dates and dispersal dates
Step 2 - Date Input/Proofing in Excel
  Create and populate Excel tables for

2a: Biograph data (include births, aborts, immigrant males)

2b: Census data

2c: Matured dates (m and f dates)

2d: Male dispersal data

2e: Female cycling data

2f: Pregnancy data

2g: Demography notes (notes on individuals from demog notes or other group notes in group binders)

2h: Other group notes (notes on individuals from 'Other Group' binders, other groups notes in study group binders, and other groups collar check sheets)

2i: Proof all data for accuracy

Step 3 – Dump and restore the Babase schema from Babase to Babase_test

Copy the Babase schema to Babase_test so trials can be run with impunity

Step 4 – empty REPSTATS, CYCSTATS, MMINTERVALS, MDINTERVALS

Step 5 – Trial uploads to Babase_test, corrections, and final uploads to Babase

5a: Export each Excel sheet to .txt file and run trail uploads to Babase_test

5b: Return to Excel to troubleshoot errors

5c: Repeat step 5b a couple hundred times ☺

5d: Save corrected Excel & .txt files and upload to Babase

Step 6 - Update the BIOGRAPH table with death and censorship information

Run SQL update command in Babase_test to change biograph values

Run SQL update command in Babase to change biograph values

Step 7 - Rebuild REPSTATS, CYCSTATS, MMINTERVALS, MDINTERVALS

Rebuild tables as database manager; then vacuum as database administrator

Step 8 – Cleanup

Pull flags off the neonatal sheets

Move neonatal sheets for infants belonging to dead moms to dead mom binder

Copy any dangling cycles to Excel files for the next update

Make notes for yourself for the next update

File handwritten sheets in plastic sheath in Demography Update Checklists binder

Finalize the update log

Make any corrections/clarifications to the Babase 2.0 protocol

Make sure completion of the update is noted on the Babase mailing list and in the Status of Babase Datasets table on the wiki homepage.

Email Duke giving the male maturedates, dispersals, ranked by dates for new males who are clearly adults (others will be assigned later @Duke), and the index of items found in notes that may not be recorded in their proper place.

Each of these eight steps is described in detail in the following sections.

Step 1: Manual data review and consolidation

1a) Jeanne's Review

Sex Skin Scoring and Infant Birthdates

Set out all the datasheets – Sex Skins, Reproductive Notes, Group Censuses, and Demography Notes – for the update period. Include one month prior to, and one month following, the three-month period of the update. Set out one or two groups at a time, separating the sex skin sheets and group censuses by month. Set out the Neonatal Notebooks as well.

Important! During the monthly proofing process, you should have drawn a line through any dates at the end of the month that don't exist in that month (remember the 30 days hath September rule). Double check that this has been done to ensure that nonexistent dates are not assigned cycling events or used in calculating birthdates.

Also mark females newly appearing on the sex skin sheets as newly added 4-year-olds.

Jeanne reviews the sex skin datasheets and marks the beginning of swelling with T or P or R (**T**urgescence or **P**uberty or **R**esumption of cycling). The tdate is marked as a P if this is the female's first cycle (indicating puberty); the tdate is marked as an R if this is the first turgescence following a pregnancy (indicating cycle resumption). Jeanne marks the end of swelling with a D or Z (deturgescence or conception). The ddate is marked with a Z if the female becomes pregnant. Jeanne uses the Neonatal sheets and the Reproductive notes to determine the birthdate, which she marks on both the sex skin sheet and the neonatal sheets. Confused? Here's a chart to visualize it:

T	=	tdate		D	=	ddate (non-concpetive)
R	=	tdate (resume cycle after ppa)		Z	=	ddate (conceptive)
P	=	tdate (puberty)				

Note: If you have followed protocols for monthly data, all infants new to this demography update period should already have their sheets flagged in the neonatals.

1b - Copy datasheet info to handwritten demography Update sheets
Review all of the information in the sex skin and census data sheets and transfer this to the handwritten demography sheets. The hyperlinks below provide guidance on how to fill in each page and section of these sheets using the data sheets from Amboseli.
Page 1   Page 3
Page 2   Page 4

1c - Send questionable pregnancies, new individual names and sexes to Team for confirmation
Jeanne will mark all the new infant birthdates on the sex skin sheets and on the neonatal observation pages. Review her scoring and all notes and questions that Jeanne may have for the Team. Also review Jeanne's notes and include any questions about cycles she may have for the Team (especially the last month of the update when new conceptions may be unclear). Include all new infant names and sexes for Team to confirm. If you are preparing an Excel file, don't forget to zip the file before you send it to Amboseli!

1d: Send your list of assigned death dates and dispersal dates around to Jeanne and Susan for approval.

Completing this step depends on having compiled page 3 of the handwritten summary sheets and this in turn depends on having a good handle on the census information. Entry of the census data is covered just below in *Step 2b of Date Input/Proofing in Excel* (Enter Census data in a table). You can enter the census data into Excel while Jeanne is scoring the sex skin information. You will need this information in order to evaluate disappearances and assign death dates and dispersal dates (for more on this refer to the link for Page 3 just above).

Step 2 - Date Input/Proofing in Excel

2a - Enter "New" Individuals in table

Follow the steps below to prepare an Excel sheet for uploading to Babase.

1. Go to C:\Database Files\ALTMANN\Babase\TEMP\REPRO.
2. Create a new folder for this update period (yya/b) if one doesn't yet exist
3. Copy the last update file (bryya/b.xls) for the group you are working with.
4. Clear all the old data
5. For each birth, immigrant male, or abort from pg 1 of the update sheets, enter data into an Excel row in the correct group worksheet
6. If individuals cannot be entered (see, Important! Below), enter them into a new workbook for the next update and put it into a new folder with that update name.
7. Once the Team has confirmed all names and sexes of new individuals, consolidate these into the worksheet "Summary".
8. Before uploading the data, be sure to proof it and indicate that you have done so in the "Proofed" comment box. If there's no date in the box, the data have not yet been proofed.

Important! New individual information is consolidated into a "Summary" worksheet because it is sometimes necessary to manipulate it (e.g. delete pids) in order to complete the upload. This is because the validation rules are so strict that you can actually reach the point where it's almost impossible to upload individuals in instances when a conception and birth occurs in the same update period. See the appendix section on

WHEN BIRTH, DEATH, AND CONCEPTION OCCUR IN THE SAME UPDATE PERIOD FOR FURTHER EXPLANATION OF THIS.

Important! Babase 2.0 will not allow you to enter any data that falls after the last census day for that group or individual. Therefore, there may be instances where a birth that occurred in late June or late Dec (after the last census day) will have to be entered as part of the *next* update. (The statdate in BIOGRAPH is automatically updated with the individual's last census day when you upload the census files.) This is also true for aborts or sexual cycle dates falling after the last census day.) You should have already identified such instances while filling out the Demography Update Summary Sheets. Be

aware of this and enter births or aborts into a bryya/b.xls file for the next update you will do (that way you won't forget to enter them).

Monthly census files are created and updated for each regularly monitored group. These files are stored in C:\Database Files\ALTMANN\Babase\TEMP\CENSUS. Follow the steps below to prepare Excel sheets for uploading to Babase.

1. Copy the census template from
    C:\Documents and Settings\altmann lab\Desktop\update_essentials\templates
2. Copy the column of snames from the last month of the previous update (either June or December).
3. Rearrange/add/delete animal snames so that they appear in the order listed on the hand-written census sheets.
4. Enter all the census dates for that month as column headers in the worksheet.
5. Indicate any absences for the month with a '0' (leave 'X's blank in Excel).
6. If an incomplete census must be entered (see below), use 'N' for individuals who were not marked either present or absent; these individuals will not receive a census record for this date.
7. Before uploading the data, be sure to proof it and indicate that you have done so in the "Proofed" comment box. If there's no date in the box, the data have not yet been proofed.

Incomplete Censuses
On some observation days, the field Team is only able to finish an incomplete census. This is indicated on the census datasheet next to census time and census recorder. Incomplete censuses usually just mark animals as present – they typically cannot confirm absences. However, there are also instances when incomplete census data is collected by researchers other than the regular field Team. In these cases the researcher may be more readily able to identify some group members (e.g., adults) and not others (e.g., juveniles) and may be able to provide accurate absence data for the individuals they are familiar with.

Incomplete censuses performed by the Team should be entered on the Excel sheet for that group, using 'N' to mark individuals that were not recorded as either present (X) or absent (0) on the census sheet. Babase will not record a census record for these individuals on that date. The presence or absence of these individuals will be interpolated as though the field Team was not with the group that day. This procedure should also be followed if a first absence can be confirmed by an incomplete census recorded by another observer. If only a first presence can be confirmed by incomplete census data recorded by another observer, the confirmed presence can be entered using the same protocols as followed for Demog Notes or Other Group Censuses (see 2g and 2h below).

Census Entry with Group Splits (Protocol used in Babase 1.0)

When a group splits, there should be no gap between the last day of the parent group and the beginning of the new groups. If the actual census records do not include the decided last and first days of these groups, manual census days should be added in order to avoid a gap in records.

For example, the split for Dotty's group occurred at the end of July 1999. The census records for Dotty's group ended on the 30th of July, and the census records for Omo's and Viola's groups began on the 3rd of August. There would therefore be a gap in records for all the animals in these groups from the 31st of July to the 2nd of August. To avoid this, the 31st was manually added as a census day for Dotty's and the 1st was added for Omo's and Viola's.

1. On census grids, if first or last day of group (usually 31st or 1st) is not an actual census day, then write in pencil records for that day and make a clear note. Remember you can use N for noninterpolating manual entry, rather than M for interpolating manual entry. Be sure all members are interpolated in the group up to the split date if there is more than 14 days between the last or first
2. Enter census records, including these added days.
3. Change the "origin" for the added days to "M" for manual. Use the following command, changing group # and date: replace all origin with "M" for grp = 1.20 and date = {31/07/99}.

Census Entry with Group Fissions (Protocol used in Babase 2.0)
Follow the above where needed and try to keep floaters, especially females, out of group 9 using strategically placed manual census points.

2c: Enter Matured dates in a table (m and f dates)
Copy the template from
C:\Documents and Settings\altmann lab\Desktop\update_essentials\templates
Fill in the template with handwritten values from the update sheets
Once you have proofed your entries, put the date in the "Proofed" comment box.

2d: Enter Male dispersal data in a table
Copy the template from
C:\Documents and Settings\altmann lab\Desktop\update_essentials\templates
Fill in the appropriate worksheet with the data from the handwritten update sheets for all groups.
Once you have proofed your entries, put the date in the "Proofed" comment box.

2e. Enter Cycles data in a table
Entering M, T, D Cycles to Excel
If you have been diligent, you should have already entered the "dangling" cycles from the previous update to a new update workbook into which all cycles will be entered.
(Dangling cycles are those that did not have a ddate by the last month of the update OR

those whose ddate fell after the last census date of the last month. Babase won't let you upload the cycle data that fall after the last census date because this is also the statdate in BIOGRAPH). Continue populating this cycles workbook or else create a new workbook to populate using the cyc0X_template in the demogr_essentials folder. If you are creating a new workbook, don't forget to review the previous update for any "dangling" cycles that should be included in the current update.

Fill in each group's Excel worksheet with the handwritten values you should have already copied to the update sheets. The five group worksheets reside in a single workbook (e.g., all female cycles for the 2006b demography update are in C:\Database Files\ALTMANN\Babase\TEMP\REPRO\Icyc06b). Later you will export these individual worksheets into five tab delimited text files for upload to Babase.

Proof your Excel entries against THE ORIGINAL SEX SKIN sheets. This is the best proofing method because your handwritten sheets could contain a transcribing error.

Once you have proofed and corrected any errors in Excel and the handwritten mtd cycles pages, note the date of proofing in the "Proofed" comment box.

Now do a SAVE AS and rename your workbook with the info for the next update. Include the beginnings of any dangling cycles from this update period.

2f. Enter Pregnancy Data in tables
After completing the Cycle entry and update, you are ready to work with the pregnancy data. Your first step in this process is to return to the Demography Update Summary Sheets and enter the CID for each Z or R entry on the Repro Sheet (you can ignore P entries). You will need this information to complete the steps that follow.

To enter the pregnancy data into Excel, copy the structure from a previous pregnancy table. Start entry of all Z's (pregnancies) on the repro sheet. Parity simply refers to the number portion of the PID (although rare exceptions are possible). In the conceive column, enter the CID; this is the same as the cpid in the cycpoints table and the dcpid from the mtd_cycles view. And don't enter anything for the resume cycle! Babase 2.0 will calculate these automatically for you.

Once you have proofed your entries, put the date in the "Proofed" comment box.

2g. Demography Notes Table
The primary objective for adding information to the Demography Notes Table is to capture census-related information that otherwise *would not* appear in the Census Table and the Members Table. Accordingly, the Demography Notes Table typically does not include information redundant with the study group censuses or attempt to manually interpolate an animal's placement in Group 9.0 ("unknown group").

General guidelines for entering records in the Demography Notes Table: Each record in the Demography Notes Table contains information linking a named individual with a group ID on a specified date. Upon update, the Demography Notes Table adds records directly to the Census Table and, consequently, affects how the Members Table is generated.

The three main sources of information usually referenced when entering records into the Demography Notes Table are:

1. Field demography notes in each of the study group notebooks
2. Other group notes in each of the study group notebooks
3. Other groups collar checks when the baboon's group is noted.

It is important to keep in mind that not all notes appearing in the sources above merit a record in the Demography Notes Table. For example, the following types of field notes are *not relevant* to the Demography Notes Table:

• Field demography notes providing general comments about the study group or miscellaneous occurrences in the field	*Not relevant because records in the Demography Notes Table are specific to a single, named individual*
• Non-study group censuses that record general information on the count of the different sexes or age classes present (i.e. sub-adult male, brown infant, or adult female)	*Not relevant because records in the Demography Notes Table are specific to a single, named individual*
• Field demography notes providing a verbal descriptive of the field census data (i.e. WAM with Omo's group today or ORI not seen in Linda's group today)	*Not relevant because the information is automatically captured in the study-group census*

With regard to record entry, remember that multiple notes pertaining to a particular individual on a given date should be combined into a single entry. Notes referring to more than one individual must appear separately for each individual (i.e. one entry per individual).

Some examples of commonly entered records in the Demography Table Notes follow. Note that this list is by no means exhaustive and record entry should not be limited solely to these scenarios.

"Alone" observations – Include a record in the Demography Notes Table for a day an animal is observed alone (Group 10.0). For example, "Liberty was observed alone today."

Other group notes – Include a record in the Demography Notes Table for any references in the "Other Group Notes" section of the monitoring notebooks that identify a named baboon. For example, "OGN: Godot seen in Ositeti's group."

Visits – Include a record in the Demography Notes Table for a day an animal "visits" a group. Observations that qualify as visits include cases where the individual was present in a group either (1) after the group had already been censused that day or (2) the individual was already censused as present in another group that same day. In both cases, the baboon will not be listed as present in the Members and Census tables in the group it visited. For example, "Liberty visited Omo's group today but was censused in Viola's group." Floaters moving between fission products fall into this category.

Death information – Information concerning the cause and timing of death for baboons with dcauses other than presumed predation or unknown should be included in a demography note on the assigned date of death (i.e., the baboon's statdate). The baboon's census point for that day should be a manual census point if the baboon was not seen that day prior to dying.

NOTE: There is now a draft deaths table in babase_pending to which these notes are also added, along with other information, such as the mother's condition if the deceased is an infant or the infant's condition if the deceased is a mother, confidence in the death and the date of death, whether the dead was witnessed or a corpse found, etc. Thus far all deaths from 2011 onward are in place and those for Linda's and Weaver's group have been backfilled along with the demography note backfill. Deaths in other groups will be backfilled as demography note backfill is completed for each group. Since the deaths table is still in draft form and incomplete, I will continue to include demography notes on the statdate for animals with information beyond having gone missing.

2h. Other Group Census Information

Other group census information is captured in a format similar to the demography notes files created for each of the study groups. This can be somewhat confusing as, by general convention, the other group information are referred to as *censuses* yet the data are entered and updated following the protocols used for study-group *demog notes*.

The files for other group census information are saved in a subfolder of the main census folder in C:\Database Files\ALTMANN\Babase\TEMP\CENSUS. The subfolder name is called "otheryy" where yy equals the last two digits of the year; each file within the folder represents all census information for the entire six-month demography time period (all groups/months entered in one file). File names are Gdyya/b/c/d with a/b/c/d representing the demography periods (again, yy equals the last two digits of the year).

1. Copy the Excel file structure from a previous "other group" census file as a template.

2. Look through other group monitoring binders for any censuses done within the demography update time period. Because these groups are not censused with the regularity of the focus study groups, many individuals are not named or recognized by the field Team. Record only positive IDs on named individuals. For example, keep in mind that "SAM" refers to "sub-adult male" and "BRI" refers to "brown infant" – these are not snames in reference to particular baboons!
   a. Include named individuals as present.
   b. Do not include any information about absences unless the quality of the census is good.
3. Once you've proofed your entry, put the date in the "Proofed" comment box.
4. Keep an eye out for new infants and female reproductive information for females in recently dropped groups. Assign nontraditional snames to new infants and put any solid reproductive information, including estimated conception and birth dates for the new infants, into cycgaps.

Step 3 – Dump and restore the babase schema from babase to babase_test

Copy the babase schema to babase_test so trials can be run with impunity

THIS SECTION REQUIRES UPDATING TOO.

The following directions are taken from the WIKI, Sept 3, 2008.

The babase-copy-babase-schema Unix program copies the entire babase schema, including data, table definitions, validation, and everything else, from one database to another. All existing data, table definitions, validation, etc is deleted from the babase schema in the target database. You must be logged in to papio and at the Unix prompt to run the program.

The following example copies the babase schema from the babase database to the babase_test database. The first database (`babase`) is the database to copy from, the second (`babase_test`) the database to copy into.

```
PGPASSWORD='*********' babase-copy-babase-schema babase babase_test
```

```
(yes, you do need to put the single quotes in there to make it run).
```

Once you hit enter, it will apparently do nothing for 20 seconds or so, but it is working. You should begin seeing the following messages pop up on the screen as it works.

```
SET
DROP SCHEMA
SET
GRANT
```

The operation is complete when you see your prompt again
`e.g., [kfenn@papio ~]$`

Step 4 – Empty REPSTATS, CYCSTATS, MMINTERVALS, MDINTERVALS

Empty the contents of these tables. In order to do this, you must log in as admin. You cannot do it using your regular login.

When you are done, VACUUM the tables to see that they are really emptied. If you have not emptied the tables properly, when you try to upload cycles information you will get nasty messages like:

```
ERROR: update or delete on "cycpoints" violates foreign key constraint
"Cpids on CYCPOINTS" on "cycstats" DETAIL: Key (cpid)=(54220) is still
referenced from table "cycstats". CONTEXT: SQL statement "DELETE FROM
cycpoints WHERE cpid = $1 " PL/pgSQL function "_remove_autom" line 56
at SQL statement SQL statement "SELECT _remove_autom( $1 , $2 , $3 )"
PL/pgSQL function "cycpoints_func" line 130 at perform SQL statement
"UPDATE cycpoints SET cpid = cpid WHERE cpid = $1 " PL/pgSQL function
"pregs_func" line 165 at SQL statement
```

These tables do interpolation that may be in conflict with the new (true) data you want to upload. So just empty them, otherwise you won't be able to do much with the cycles data. You will rebuild them again manually at the end.

Step 5 – Trial uploads to Babase_test, corrections, and final uploads to Babase

The data files are uploaded in the following order.
1 - Copy the Babase schema from BABASE to BABASE_TEST
2 – Empty REPSTATS, CYCSTATS, MMINTERVALS, MDINTERVALS
3 – Uploaded any missing conceptions from the end of the last update to PREGS and/or delete any false conceptions from the last update
4 – Upload the births file to BIOGRAPH
5 – Upload census files to CENSUS
6 – Upload demog notes to DEMOG_CENSUS (view)
7 – Upload the OTHER GROUPS demog notes to DEMOG_CENSUS
8 – Upload male and female maturity dates to MATUREDATES
9 – Upload dispersals to DISPERSEDATES (table)
10 – Upload cycles to MTD_CYCLES (view)

11 – Go get all those dcpid numbers for the pregs table upload

12 – Upload pregs table to PREGS

13 – Kill the dead and censor the long-absent baboons in BIOGRAPH and give manual census points on date of exit from the population

14 – Rebuild the reproductive tables

15 – Upload manual weather data to MIN_MAXS

16 – Upload WeatherHawk files to babase_pending

17 – Upload wounds and pathologies files to babase_pending

18 – Upload neonatals to babase_pending

19 – Upload subgroup notes to babase_pending

20 – Upload intergroup encounters to babase_pending

21 – Upload raw hybridity scores (and hybridmorpho where applicable) to babase_pending

22 – Upload SEXSKINS and PCSKINS (if any are ready)

23 – Move neonatal sheets of dead mothers, write death and dispersal dates on census sheets, update the babase status table on the wiki, and email indexable items from demog and other groups notes to Duke

Step 6. Update deaths in BIOGRAPH (kill the dead baboons) and censor individuals not seen for two or more quarters.

Data on the death of an individual is derived from census sheets and entered in BIOGRAPH. Unlike prior steps, you do not prepare a "deaths" table to upload. This is one of the few cases where standard protocol is to manually edit the data in the statdate, status, dcause, and dcauseconfidence columns of BIOGRAPH.

Status – 0 = alive, 1 = dead, 2 = censored (animal-driven), 3 = censored (observer-driven)

Statdate – This is the date of the last census record for a live individual in Babase
This is the death date for individuals with a status of '1'
This is the censor date for individuals with a status of '2' or '3'

Dcause – Refer to Appendix 2 for details on assigning dcauses

Dcauseconfidence – Also refer to Appendix 2

Individuals with a status other than 0 are considered "not alive" by Babase. Thus new data falling after the statdate cannot be entered for baboons with a status of 1, 2, or 3. New data can be entered after the statdate for individuals considered to be alive (status = 0); Babase automatically updates the statdate to reflect the newly uploaded information. For details on assigning death dates and censor dates, see the Deaths and Censorship section of Appendix 1.

Record all deaths on the demography update sheet and then manually update the pertinent BIOGRAPH fields with this data using SQL UPDATE commands. As part of the step, it is also a good time to move the neonatal sheets for the infants of newly dead females from the living mother binder to the dead mother binder. Also be sure to mark the dates of death on the census sheets where the individuals last appeared.

> Run SQL update command in babase_test to change biograph values
> Run SQL update command in babase to change biograph values

Beginning in January 2010, baboons can be censored; prior to this baboons were either alive or dead in BIOGRAPH. A baboon is censored when it has not appeared in the census records for at least a six-month period. Since individuals with a censored status in Babase are considered to be "not alive", this prevents data from accidentally being entered for individuals that were not actually seen. Censoring allows us to do this without marking the individuals dead.

Two main groups of animals are censored. The first round of censoring included a large number of individuals from groups no longer being studied (e.g., the Lodge groups). A good many of these individuals surely are dead by now but the timing and means of their deaths are entirely unknown to us; therefore we cannot mark them dead. Censoring them allows us to treat them as dead without declaring them dead. Three study groups (Kelly's, Mica's, and Snap's) were dropped in 2012 and members of these groups are also being censored as they drop off the other groups census sheets. The other main group of individuals being censored consists of males that emigrate from study groups and are not seen for a period of at least six months. Some of these males may show up in future censuses, at which time their statuses will be changed back to alive. This change will need to occur before any new census data is uploaded for those individuals. Individuals not seen for a six-month period should be assigned a status of censored at the same time that individuals known to have died are changed to dead in Babase.

Initially all censored animals were assigned a dcause of 2 but in mid-2014 we divided these two types of cases into two different censored categories with males who have left study groups, i.e., animal-driven cases, retaining a dcause of 2 while cases where the animals were no longer being monitored due to observer behavior, i.e., observer-driven cases, being reassigned to a dcause of 3.

Beginning in the second half of 2011 we also assign a manual census point (M or N) on the exit date for all baboons that die, disperse, or are censored when the date is assigned by picking a midpoint between the last date present and the first date absent so that they

are present in the correct group on their exit date. If a baboon is believed to have died and the death is assigned as other than a presumed predation, the baboon should also be given a demography note on the deathdate describing the details of the baboon's death. These demography notes may eventually be found only in the deaths table (now in babase_pending), but for now can be found in demography notes on the statdate. Demography notes describing information on deaths prior to this new protocol will be added (or moved to the correct date) during demography note backfill and dcauseconfidence assignment, which is currently in progress.

Step 7 - Rebuild REPSTATS, CYCSTATS, MMINTERVALS, MDINTERVALS

Rebuild tables as dbase manager, then vacuum as dbase admin.

```
SELECT rebuild_all_repstats();
SELECT rebuild_all_cycstats();
SELECT rebuild_all_mmintervals();
SELECT rebuild_all_mdintervals();

We used to rebuild members also, using the following command, but
members updates itself automatically so this is no longer necessary.
SELECT rebuild_all_members();
```

Step 8 – Cleanup

At the conclusion of every update it is very important to
1) Make sure you moved the offspring of dead females from the neonatal books.
2) VACUUM the tables you have updated so row counts are accurate.
3) Adjust and cleanup any additional items in need of correction.
4) Email any information Duke will need for their half of the update.

HORMONE DATA

At present, data from hormone analyses performed in the Princeton lab are *not* incorporated into the Babase database. Rather, this data – contained in two tables (FECALPREP and FecalResults) – exist separate from the project's relational database, Babase. Although the entry and management of these tables is primarily the responsibility of the lab manager (Laurence Gesquiere), the database manager does take part in saving archived copies of the data and performing routine checks of data integrity.

Archiving
The lab manager will periodically send copies of Fecal PREP and Fecal RESULTS tables to the database manager (typically, the timing depends upon the rate of lab work and database changes). These copies should be saved on the database manager's harddrive in the following directory: C:\Database Files\ALTMANN\Hormone Lab (simply replace older files when newer versions are available). At regular intervals, back-up the entire

hormone lab folder along with other subfolders in the Altmann user's directory to the external hard drive.

Integrity Checks
It is useful to query the data in Fecal PREP against relevant Babase tables on a regular basis. An ideal time to do this is immediately following each six-month demography update, though this isn't a specified step of the demography update itself. D. Onderdonk developed a set of queries that are extremely useful when querying Fecal PREP for data errors. *Refer to **Appendix 3** for the specific queries and the logic behind them.*
Any possible errors in Fecal PREP that are detected using these queries should be discussed with the lab manager – depending on the nature of the problem, either the lab manager or the database manager will correct them.

RANKING

A Brief Overview
Note that while this overview of ranking protocols generally holds true for assigning both male and female ranks, some differences in decision rules for male and female rank do exist. Refer to the Duke protocols for details on male rankings.

For the baboon groups we are working with, we assume a linear dominance hierarchy. Each member in a group has a unique numerical rank that represents its position in the hierarchy relative to all other same-gendered members of the group for a particular month (*i.e.* ranks are tabulated for groups for every month). The ranks are determined by the outcome of decided agonistic interactions and are considered to remain stable unless there is some evidence to justify a change. An individual can be beaten by any other baboon that ranks above it, but is not often beaten by an individual correctly ranked below it. If an individual is beaten by a lower ranked member of the same class, this may indicate a change in ranks.

Female ranks are entered and updated in Babase using the Ranker program for all female ranks (ALF) and adult female ranks (ADF), respectively. The Ranker program displays decided agonism data for a particular group and time period in a win/loss matrix showing the number of times every individual beats (down) or is beaten (across) in an agonistic interaction. Note that the default rank order is alphabetical and in order to select the most recent month's rank, you must specify this in the program (more detailed directions are given below). Any new members to the group are added to the bottom of the list in descending age order (when working with the female ranks, typically the only new members to a group are young infants).

Decision Rules for Assigning Ranks
It is important to keep in mind that few set rules exist for assigning ranks. Rather, by considering interactions both in the immediate month and over longer periods of time and

(to a far lesser degree) considering an individual's family line and age, you begin to get a sense of what does or does not justify a change in rank position. Some basic questions to ask yourself when considering each reversal are

1. Do the numbers for that month justify giving credit?
2. Do the numbers looking forward justify giving credit?
3. Does giving credit create new reversals moving forward?
4. Is the reversal offset that same month?
5. Are there baboons in between the reversal for which you create new reversals?
6. What does the big picture look like?

The following, then, are the general steps to follow when assigning ranks – the details are best sorted out during training or by referring to old matrices and considering the decisions made in the past.

1. Make sure that both Psion and decided agonisms have been updated through the rank time period (and therefore the demography update will also have been completed).
2. You may want to make a diagram of the familial relationships between the females who appear in the ranks for that year. This will provide extra context when later considering certain rank decisions. Also, look at the last rank changes made in the previous months to get an idea of who may still be adjusting positions within the group. Currently female lines are diagrammed in PowerPoint.
3. Two options exist for tracking the changes you suggest, each month by month. One feasible tactic is to continue reviewing all monthly matrices *without* making any changes to the ranks as you progress. As a result, all matrices will reflect the rank order of the last month entered into Babase. While this option may save time in some situations (particularly when working with smaller groups), it can get overwhelming when tackling the ranks of larger study groups. In these cases, it may be useful to reprint the matrices for each month, adjusting for rank changes as you go. Be certain *not* to save the ranks as you work through the months – *only save the display*. Also, keep in mind that you will have to repeatedly enter rank changes for every month after a change has been assigned (since the ranks are *not* being saved in the database).

Updating Ranks
Once you have considered all months/groups, go over the matrices and decisions with Jeanne. After finalizing the ranks with Jeanne, you are ready to return to the ranker program and save the rankings. Immediately after saving the all female ranks (ALF) for a group-month, save the adult female ranks (ADF) for that same group-month by using the subrank program.

Protocol for Assigning Female Ranks
Appended by T. Fenn April 2009

Before you can do the ranks for a given year, Duke must first load the agonisms for that period

Using Ranker for Babase 2.0
Step 1: Connect to the VPN:
Step 2: Launch *Ranker*
Step 3: Create 15 ALF Matrices for each group
Step 4: Manipulate the rank order
Step 5: Output the Matrices to .txt

CONVENTIONS FOR NAMING MATRICES
READING THE RANKING MATRICES
IDENTIFYING ALL RANK REVERSALS
DECISION RULES FOR ASSIGNING RANKS
MANIPULATING THE RANK ORDER
RANKING ADFS

<mark>History and Overview</mark>
With Babase 1.0, K. Pinc developed a *Ranker* program for FoxPro. This will be referred to as *old Ranker* or *Ranker 1.0*. Refer to **FRank_protocol.doc** or the section on ranks within **Data Management Protocols August 2006.doc** if you have any interest in execution of the older program. *Ranker 2.0* is the new version of the ranker program that is integrated with Babase 2.0. It was developed at Duke by undergraduate Tylor Brock and updated by his advisor, Jun Yang. It is housed on the papio server.

At Princeton, we do All Female Ranks (ALF) and Adult Female Ranks (ADF). Duke is responsible for the male ranks. A description of the male rank protocol should be available on the Wiki. Always create the ALF ranks first. In Babase 1.0 there was also a program called "SubRanker" that extracts the ADF from the ALF ranks. This is not a separate program in Ranker 2.0. Rather, you just select ADF from the dropdown menu at the beginning to seed the program for this.

The database manager creates the agonism matrix outputs (using RANKER) and does a first pass on rank changes (using your brain and some of the guidelines listed below). The suggested rank changes then go to Jeanne for review. She will provide comments and request additional queries. Follow up with these and suggest appropriate changes based on the results, then meet with Jeanne to do a second pass on the rank assignments. There will be several versions of the agonism matrices before Jeanne settles on a 'final' ranking for each month. CAFEFULLY keep each version in separate folders within the group + year folders and be highly organized with the outputs so you don't confuse yourself or Jeanne. For example, the pathway for Weaver's 2007 ranks would be:

C:\Database Files\ALTMANN\Babase\TEMP\Ranks\Weaver\2007. Inside this folder would be several more folders titled **v1**, **v2**, **v3** to indicate each new run of the matrices. Now is a good time to get those v1 folders set up in a folder for the year you are doing.

Examples given below are for 2006 matrices but dates should be adjusted accordingly for new years.

Prerequisites for using *Ranker for Babase2.0*
The female rank assignments are done by year. *Ranker 2.0* creates interaction matrices by drawing agonism data directly from Babase. Therefore, all agonism data for the entire calendar year must first be uploaded to Babase 2.0. (This is part of Duke's 6 month update.) The agonistic interactions are updated within PARTS and INTERACT but the easiest way to review the raw data is through the ACTOR_ACTEES view.

From Princeton, you must run the *VPN* program to access and run *Ranker*. The protocol to download and setup *VPN* is available on the Babase Wiki. The directions below assume you have done this already.

Step 1: Connect to the VPN
In the bottom toolbar right click on the VPN logo and select "Connect". These dialogue boxes will popup.
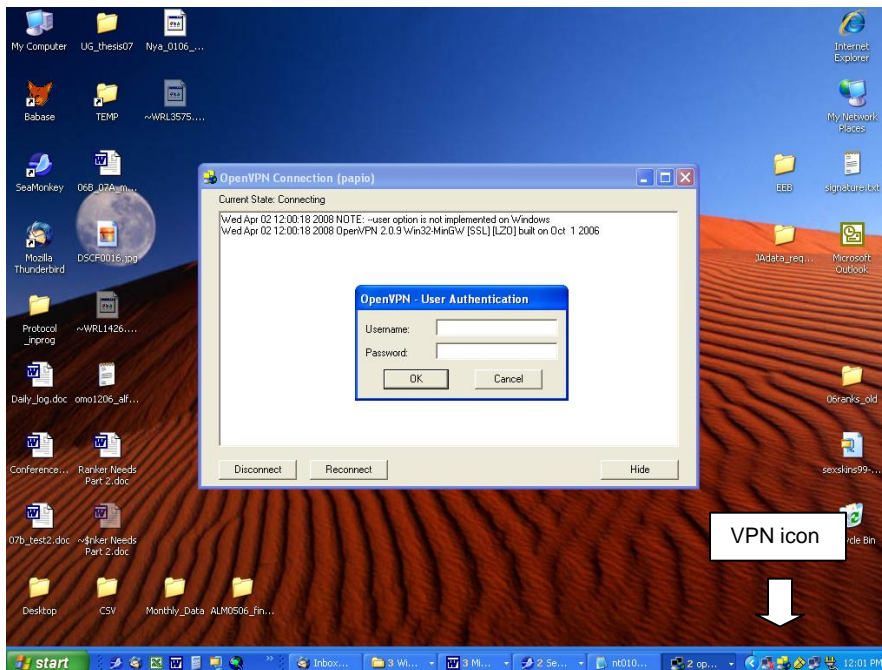
Type in your papio login and password and hit OK. The box will fill with line commands over the next few seconds and then you will get a brief message over the VPN icon in the bottom toolbar saying that you are now connected to papio.

If you are uncertain about the connection you can "ping" the server.
Go to Start > Run and type
`ping papio-vpn.biology.duke.edu`
It should 'ping' the network and give you a line beginning with "Reply" three times if the connection is good.

Note on Windows accounts:
K. Pinc has instructions for running the VPN as an admin from a limited account on a Windows machine, but I still cannot make this work. Therefore, when I need to work with Ranker 2.0 extensively for 1-2 weeks, I give myself admin privileges on my Princeton computer and remove them once I have finished. This works fine and doesn't make the Princeton IT folks mad (they don't like us to work in admin accounts constantly).

Step 2: Launch *Ranker*
Double click the **ranker** icon, located on the desktop of the database manager's computer.

You will see this box

Select
Create/load ranking from Babase

You will see this screen

Type your Babase (PhpPgAdmin) username and password and hit Connect. (Leave the ssl box unchecked).

Now you see this screen.

Step 3: Create ALF Matrices for Each Group

Create a total of 15 matrices *for each group*: one matrix for each month (12 matrices total), and three additional cumulative matrices (explained below). To do this, first generate, then output and name each individual matrix in text format. Do the ALF ranks for the individual months first. Follow the steps below.

- Choose the ALF option in the first box.
- Select the group to be ranked in the second box.
- Put the range of dates of interest in the 3rd and 4th boxes (if this is your first matrix, do January (e.g.,. start date 2008-01-01, end date 2008-01-31)
- Leave the default interaction class as 'A'
- Repeat this range of dates in the 5th and 6th boxes.
- Hit Rank

Because there is no existing rank in Babase for January of 2008, you will get a message like the following. Answer YES, you want to use the rank from the previous month. (You should have a December rank order. If it's missing, you never input it from the last ranks update).
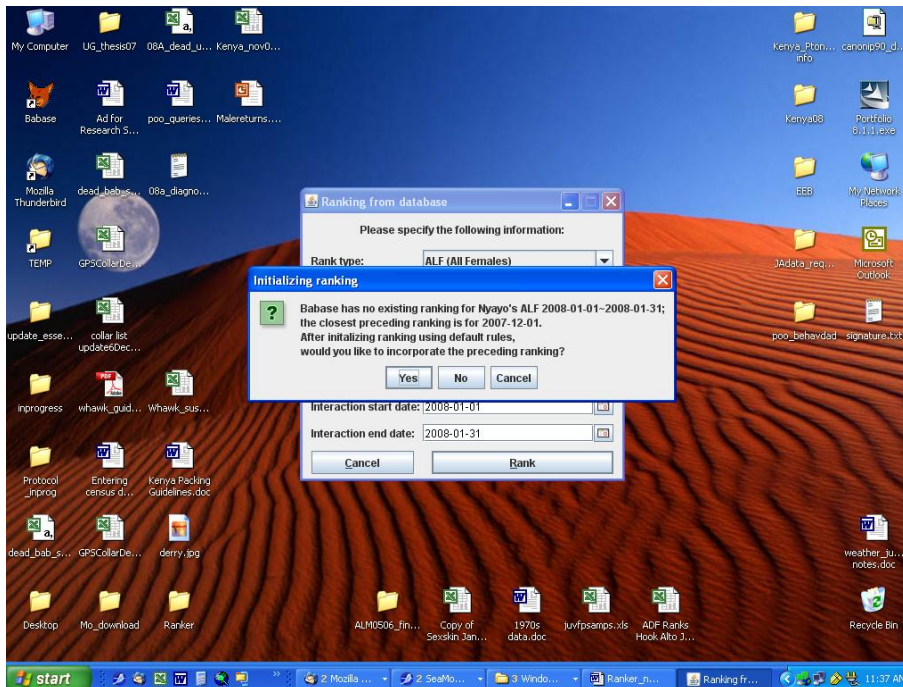
Then you get a screen that shows the proposed rank order using the most recent month's data (in this case, December).
Say YES to this also. The default is to create the ranks alphabetically and that order is almost never desirable or sensible.
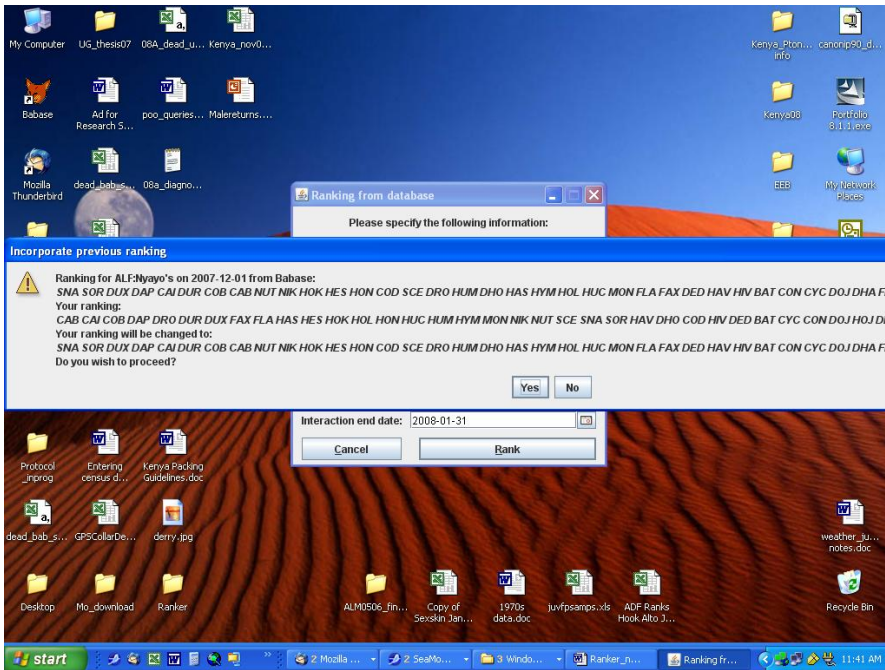
This is the matrix *Ranker* generates given those parameters.

Note that Ranker has generates a list of snames using the December order. For the first matrix run (i.e. the first 15 files for each group) use the rank order from December of the previous year. This allows you (and Jeanne) to evaluate the agonistic interactions under the assumption that no rank changes need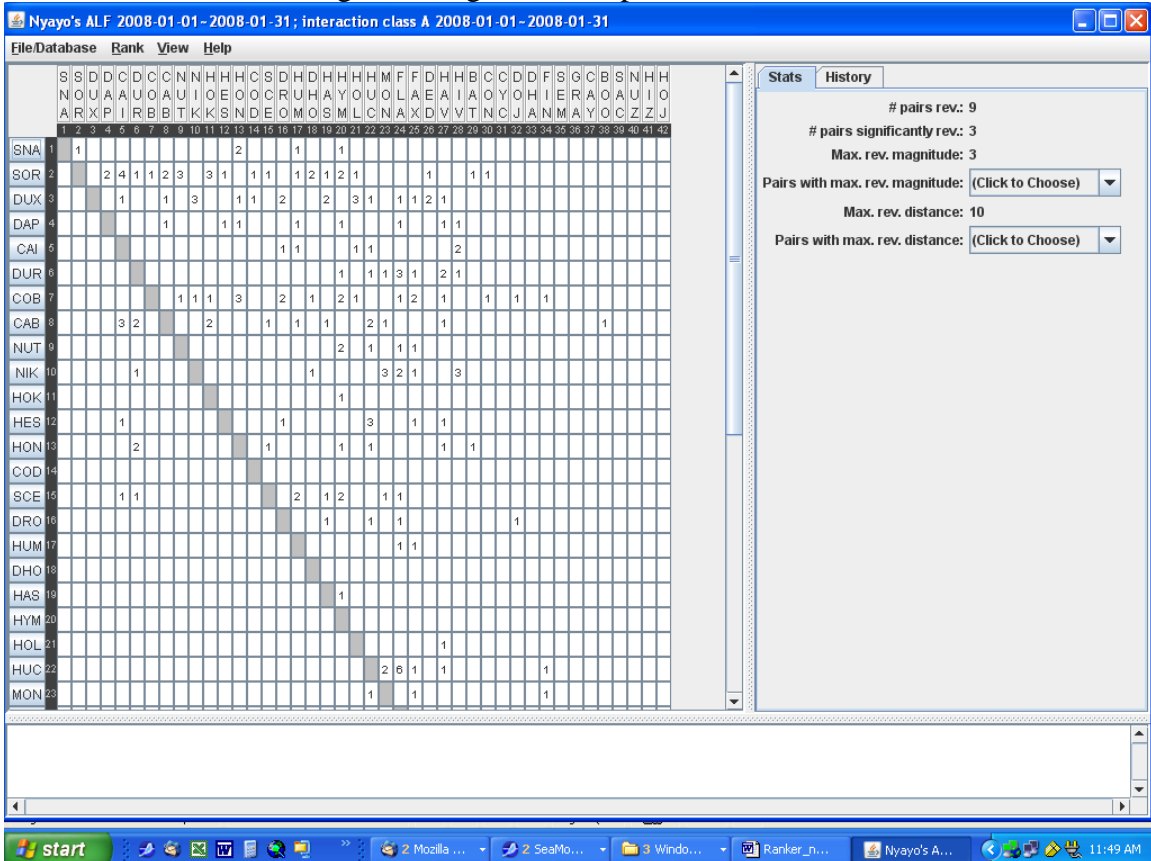 to occur…an assumption that is almost always wrong.   However, this is the best way to highlight the changes that *do* need to occur. The reasons behind this will be discussed in the section <u>Decision Rules for Assigning Ranks.</u>

Once the monthly matrices are created, you can also make the cumulative matrices. Repeat steps 2 - 5 above to create the cumulative matrices, but expand the date range. Create one matrix to show all agonistic interactions for the entire year (Jan-December). Create two 'half year' matrices that show agonistic relationships from Jan-June and July-Dec. Use these date ranges for each cumulative matrix.

	Full Year	Half Year (06a)	Half Year (06b)
Start Date	2006-01-01	2006-01-01	2006-07-01
End Date	2006-12-31	2006-06-30	2006-12-31

Be sure to know your "30 days hath September" rules so you don't accidentally exclude the 31st day in some months (January, March, May, July, August, October, December). Do not worry about putting in 31 days when there are only 30 – ranker will give you an error message saying 'invalid date' if you do. Follow the naming conventions and tag each file with 'v1' so you know when you ran them. Be sure to save them to the correct folders!

Step 4: Manipulate the rank order
The Ranker program allows you to drag and drop individuals to different positions in the matrix to see the proposed outcome. In your first run, you will not need to manipulate the rank order. Simply print out the matrix using the December order for all months and skip onto the next step. In later runs (especially if you are working with a large, confusing group) may choose to show your proposed changes each month in the printout and then seed the following month's matrix using the new rank order you proposed. Just be sure that if you choose to do this, you keep careful track of the output matrices that you have manipulated. Don't confuse the true rank order with your suggestions. Make it clear to Jeanne when you give her the set of matrices for each group just how you decided to do them, showing proposed changes month by month or showing only the rank order from the previous December.

Step 5: Convert the Matrices to .txt

Once you have created a matrix, output the results to a .txt file so it can be printed, all reversals marked and reviewed, suggestions made, and ideas exchanged with Jeanne.

Go to File/Database > Print Ranker State to Text File
Use the popup dialogue box to navigate to your v1 folder in the correct group folder. Name your matrix according to the file naming rules (just below) and be sure to tag this run as "v1" so you know which run your printout came from (ex. nf0108_v1).

You can then go to File/Database > Close Current Ranking
You will get a message saying "This rank has not been saved to Babase are you sure you want to close it?" Yes you do.
Do not save any ranking to Babase until it is the absolute final version. You can print whatever matrices you want as text but keep them well organized and do not save into Babase until you have final approval on everything.

Once you close the ranking, you go to a screen that looks like this:

After you have exported all the groups monthly and cumulative files from Ranker, go to your folder and open them in wordpad or notepad. You must save them as text if necessary if your computer doesn't automatically read them that way. When you first save the file from the RANKER window to .txt, if it doesn't show up in the ranker window in the directory, then you have to go into your folder from the explorer, open file with the funny windows icon in wordpad and then save it as a .txt file. Then it will show up in the RANKER directory when you are trying to save the next file and sometimes the subsequent files will just show up as text without this extra step.

To print them out, change the setup to Landscape and change the font size to 8 point and/or adjust margins as needed for larger groups like Nyayo's, Viola's, and Weaver's. You need to see the entire matrix on a single page. (The descriptive statistics can be on the back.)

Click the top option Create/Load Ranking from Babase and start the process all over again for the February to December data (using Dec of the previous year as the rank order). Be sure to output them to the correct folders.

Each matrix name will have the first letter of the group's name (except for Linda's which uses the letter "I"). This is followed by a 't' to indicate this is a temporary matrix. (Use an 'f' when the matrix is final and has been approved by Jeanne.) End with a numeric value to indicate the month and year the matrix covers.
Example: a temporary matrix for Weaver's Jan 2006 agonism is **wt0106**

The convention for the cumulative matrices is a little more complicated
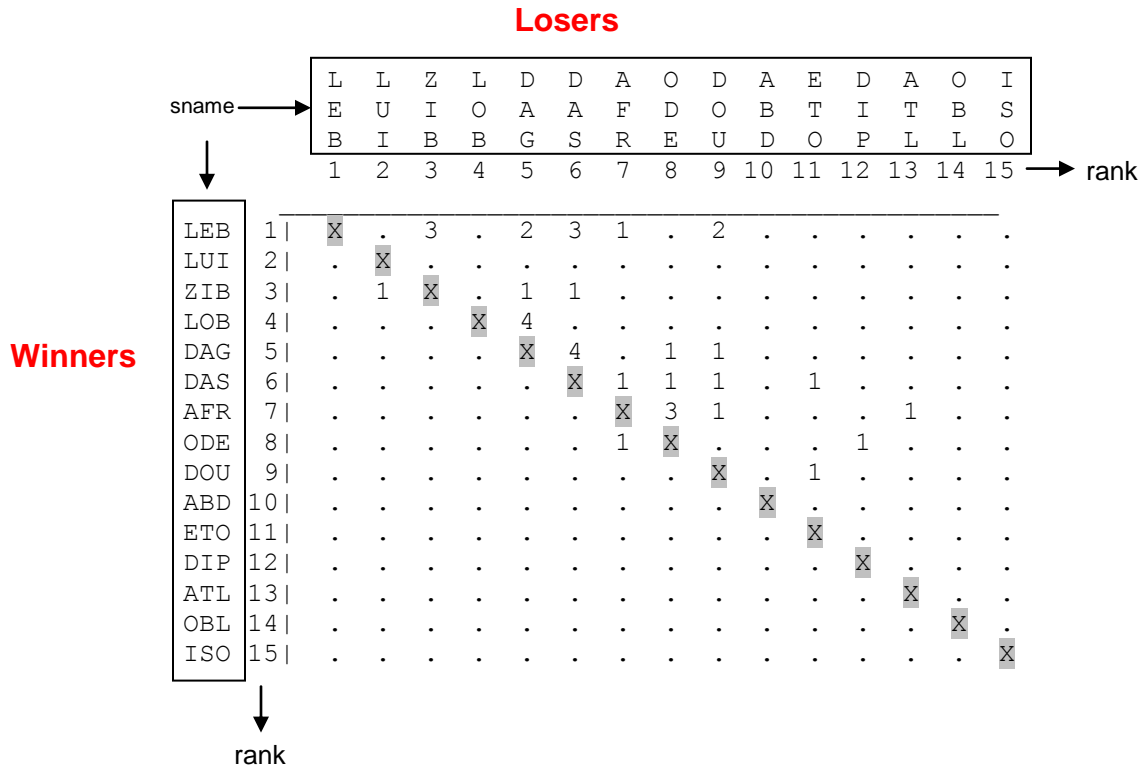**wt010606** – covers January (**01**) to June (**06**) of 2006 (**06**)
wt071206 – covers July (**07**) to December (**12**) of 2006 (**06**)
wt011206 – covers January (**01**) to December (**12**) of 2006 (**06**)

Below is a sample output of female agonistic interactions and ranks for Omo's group. The female snames are listed vertically and horizontally, as are their ranks. The diagonal line of "X"s separates those individuals who won or lost to a higher ranked individual. If ranks are correct, all lower ranked individuals should lose to higher ranked ones and there should be NO numbers below the "X" diagonal.

Losers

```
                      L  L  Z  L  D  D  A  O  D  A  E  D  A  O  I
        sname ──────→ E  U  I  O  A  A  F  D  O  B  T  I  T  B  S
                      B  I  B  B  G  S  R  E  U  D  O  P  L  L  O
                      1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 ──→ rank

          LEB  1|  X  .  3  .  2  3  1  .  2  .  .  .  .  .  .
          LUI  2|  .  X  .  .  .  .  .  .  .  .  .  .  .  .  .
          ZIB  3|  .  1  X  .  1  1  .  .  .  .  .  .  .  .  .
          LOB  4|  .  .  .  X  4  .  .  .  .  .  .  .  .  .  .
Winners   DAG  5|  .  .  .  .  X  4  .  1  1  .  .  .  .  .  .
          DAS  6|  .  .  .  .  .  X  1  1  1  .  1  .  .  .  .
          AFR  7|  .  .  .  .  .  .  X  3  1  .  .  .  1  .  .
          ODE  8|  .  .  .  .  .  .  1  X  .  .  .  1  .  .  .
          DOU  9|  .  .  .  .  .  .  .  .  X  .  1  .  .  .  .
          ABD 10|  .  .  .  .  .  .  .  .  .  X  .  .  .  .  .
          ETO 11|  .  .  .  .  .  .  .  .  .  .  X  .  .  .  .
          DIP 12|  .  .  .  .  .  .  .  .  .  .  .  X  .  .  .
          ATL 13|  .  .  .  .  .  .  .  .  .  .  .  .  X  .  .
          OBL 14|  .  .  .  .  .  .  .  .  .  .  .  .  .  X  .
          ISO 15|  .  .  .  .  .  .  .  .  .  .  .  .  .  .  X

               rank
```

Now that you have created the 15 matrices specified, review the 'reversals' between interacting pairs. Anywhere a number appears below (i.e. to the left) of the diagonal line, this is a 'rank reversal". This is where a lower ranked individual won over a higher ranked individual. The following screenshot shows rank reversals in a matrix from the old FoxPro Ranker, but it illustrates the point.

```
Microsoft Visual FoxPro
File  Edit  View  Tools  Program  Window  Help

Interaction Matrix: Actor is row, Actee is column
          P  P  P  W  W  P  W  W  W  W  W  L  W  L  L  L  L  L  L  K  L  K  K  L  L  W  L  L
          E  A  O  E  H  E  H  Y  O  E  I  A  E  E  A  O  I  I  Y  O  E  U  O  I  U  O  A  Y  U
          B  I  K  A  E  M  O  N  B  N  R  S  D  I  Z  C  Z  M  E  X  L  P  L  W  R  M  W  M  N
          1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
PEB  1|\_ .  .  .  .  1  .  .  .  .  .  .  .  .  .  .  .  .  1  .  .  .  .  .  .  .  .  .  .
PAI  2| . \_ .  .  .  1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  1  .  .  .
POK  3| .  . \_ .  .  .  4  2  .  .  2  .  2  .  .  1  .  .  .  .  .  .  .  .  1  .  .  1  .
WEA  4| .  .  . \_ .  .  .  .  .  .  1  .  .  .  2  .  .  .  .  .  1  .  .  .  .  1  .  .
WHE  5| .  .  .  . \_ .  .  .  .  .  .  .  .  2  .  .  .  .  .  .  1  .  .  .  .  .  .  .
PEM  6| .  .  .  ① . \_ .  .  .  .  1  .  .  .  .  .  .  .  .  .  .  .  1  .  .  .  .  .
WHO  7| .  .  .  .  .  . \_ .  .  .  .  2  .  .  2  .  .  .  .  1  .  .  .  .  .  .  .  .
WYN  8| .  .  .  .  .  .  . \_ 1  .  .  .  .  1  .  1  .  .  .  1  .  .  .  .  .  1  .  .
WOB  9| .  .  .  .  .  .  .  . \_ .  .  .  .  .  .  .  .  .  .  1  .  .  .  .  .  .  .  .
WEN 10| .  .  .  .  .  .  .  .  . \_ .  .  .  .  .  1  .  .  .  .  .  .  .  .  .  .  .  .
WIR 11| .  .  .  .  .  .  .  .  .  . \_ .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
LAS 12| .  .  .  .  .  .  .  .  .  .  . \_ .  .  1  .  .  .  1  1  .  .  .  .  .  .  .  .
WED 13| .  .  .  .  .  .  .  .  ① .  .  . \_ .  .  .  .  .  1  .  1  1  .  2  1  .  .  .
LEI 14| .  .  .  .  .  .  .  .  .  .  .  .  . \_ .  .  .  .  .  1  .  .  .  .  .  .  .  .
LAZ 15| .  .  .  .  .  .  .  .  .  .  .  .  .  . \_ .  .  .  .  .  .  .  .  .  .  .  .  .
LOC 16| .  .  .  .  .  .  .  .  .  .  .  .  .  .  . \_ .  .  .  1  .  .  .  .  .  .  .  1
LIZ 17| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . \_ .  .  1  2  1  .  .  .  .  1  .
LIM 18| .  .  .  .  .  .  .  .  ① .  .  .  .  .  .  .  . \_ .  .  .  .  .  .  .  .  .  .
LYE 19| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . \_ .  1  .  .  1  .  .  .  1
LOX 20| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . \_ .  .  .  .  .  .  .  .
KEL 21| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . \_ .  1  .  .  .  .  .
LUP 22| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  . \_ .  .  .  .  .  .
KOL 23| .  .  .  .  .  .  .  .  ① ...
```

Recalling the Winners / Losers designation from the first matrix, the following interaction would be read as:

"PEM won over WEA one time."
Or, if you want to read it in reverse:
"WEA lost to PEM one time."

This is interesting, but it only tells half the story. Suppose we want to know how many times WEA *defeated* PEM? Because each individual's sname shows up twice on the matrix you can see the total number of wins and losses for that interacting pair, as illustrated below for three different pairs.

```
 Microsoft Visual FoxPro
 File  Edit  View  Tools  Program  Window  Help

 Interaction Matrix: Actor is row, Actee is column
          P  P  P  W  W  P  W  W  W  W  W  L  W  L  L  L  L  L  L  K  L  K  K  L  L  W  L  L
          E  A  O  E  H  E  H  Y  O  E  I  A  E  E  A  O  I  Y  O  E  U  O  I  U  O  A  Y  U
          B  I  K  A  E  M  O  N  B  N  R  S  D  I  Z  C  Z  M  E  X  L  P  L  W  R  M  W  M  N
          1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
 PEB  1|\_  .  .  .  .  1  .  .  .  .  .  .  .  .  .  .  .  .  1  .  .  .  .  .  .  .  .  .  .
 PAI  2| .  \_ .  .  .  1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  1  .  .  .  .
 POK  3| .  .  \_ .  .  .  4  2  .  .  2  .  2  .  .  1  .  .  .  .  .  .  .  1  .  .  1  .
 WEA  4| .  .  .  \_ .  (.) .  .  .  .  .  1  .  .  .  2  .  .  .  .  .  1  .  .  .  1  .  .
 WHE  5| .  .  .  .  \_ .  .  .  .  .  .  .  .  2  .  .  .  .  .  1  .  .  .  .  .  .  .
 PEM  6| .  .  .  (1) . \_  .  .  .  .  .  1  .  .  .  .  .  .  .  .  .  1  .  .  .  .  .  .
 WHO  7| .  .  .  .  .  .  \_ .  .  .  .  2  .  .  2  .  .  .  1  .  .  .  .  .  .  .  .
 WYN  8| .  .  .  .  .  .  .  \_ 1  .  .  .  .  1  .  1  .  .  1  .  .  .  .  .  .  1  .
 WOB  9| .  .  .  .  .  .  .  .  \_ .  .  .  (.) .  .  .  .  1  .  .  .  .  .  .  .  .
 WEN 10| .  .  .  .  .  .  .  .  .  \_ .  .  .  .  .  1  .  .  .  .  .  .  .  .  .  .
 WIR 11| .  .  .  .  .  .  .  .  .  .  \_ .  .  .  1  .  .  .  .  .  .  .  .  .  .  .
 LAS 12| .  .  .  .  .  .  .  .  .  .  .  \_ .  .  1  .  .  (.) 1  1  .  .  .  .  .  .
 WED 13| .  .  .  .  .  .  .  (1) .  .  .  .  \_ .  .  .  .  .  1  .  1  1  .  2  1  .  .
 LEI 14| .  .  .  .  .  .  .  .  .  .  .  .  .  \_ .  .  .  .  1  .  .  .  .  .  .
 LAZ 15| .  .  .  .  .  .  .  .  .  .  .  .  .  .  \_ .  .  .  .  .  .  .  .  .  .
 LOC 16| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  \_ .  .  1  .  .  .  .  .  .  1
 LIZ 17| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  \_ .  1  2  1  .  .  .  1  .
 LIM 18| .  .  .  .  .  .  .  .  (1) .  .  .  .  .  .  .  .  \_ .  .  .  .  .  .
 LYE 19| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  \_ .  1  .  .  1  .  .  1
 LOX 20| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  \_ .  .  .  .
 KEL 21| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  \_ .  1  .  .
 LUP 22| .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  \_
 LOI 23|                                  1                              \
```

The first pairing would be read:
PEM won over WEA 1 time.
WEA won over PEM 0 times.

Our shorthand for this is:
PEM>WEA 1:0

After printing all 12 monthly matrices, go through them and circle each reversal and find the corresponding losses for that interacting pair. Write every single reversal out using the shorthand above. In the past we've put each one on a little yellow stickie but recently transitioned to typing everything up in Excel so it's neater and can be copied easily.

Take the cumulative matrices and find the same pair in the Jan-June matrix and record the win:loss results. Record the same information from the July-December time period. Below each reversal, put 06A and 06B and note the results.

The final results would hypothetically look like this:
PEM>WEA 1:0
06A 3:2
06B 2:0

These values allow you (and Jeanne) to look both forward and backward in time to understand the context in which the particular reversal occurred.

Once you have compiled all the win:loss information for every month, review the next section Decision Rules for Assigning Ranks and determine if there is any reason to move an individual up or down in rank.

A shortcut for gleaning information from the cumulative matrices:
T. Fenn prepared a couple of macros that can be helpful in pulling together the cumulative data so you don't go cross-eyed trying to follow the diagonal lines between interaction in the matrices. The macros are in the file:
C:\Database Files\ALTMANN\Babase\TEMP\Ranks\ADF_ALF_07a

Open this file
Open a new .xls workbook
Create worksheets for each cumulative matrix for each group (Jan- Jun and Jul-Dec for each group: 10 total)
Import the .txt files with the cumulative matrices to each worksheet (you can add column lines in the Import wizard if the separations between columns aren't correct).
Run the sequence of macros described in the "macro" worksheet of the 07 file you opened.
Now you have the cumulative matrix reversals shown below the diagonal "X" with the win/loss pair displayed side by side. She typically did a little more reformatting on these – font size of 7, add borders between pairs, and resize the columns, delete any extra text around the matrix - then print them out as a quick and easy summary of what each individual did each month.

Another way to prepare the "Stickies"
Go to C:\Database Files\ALTMANN\Babase\TEMP\Ranks\ALF_ADF_08 and copy the template. This is an alternative form to doing the stickies by hand. It's also useful for creating a pivot table that summarizes the months when reversals occurred between a specific pair.

Copy the template into a worksheet for each group. Fill in the boxes of the template showing the reversals in the shorthand given above. Complete the A and B cumulative reversal summary. Then run a pivot table with winners on the left, months on the left, losers across the top, and SUM of count column in the data box. Check the totals against the totals in the cumulative matrix to make sure you didn't forget to input any pair reversals. Proof all your reversals one more time, along with the A and B values you recorded. Once you are sure the data is correct, you can begin deciding whether the winner should get credit and move up in rank.

One thing to check: look at any of the reversals where the numbers are rather high for the winner and low for the loser, especially in the first 6 months. Go back to the previous 6 months and see if you gave credit where perhaps you shouldn't have.

Note that while this overview of ranking protocols generally holds true for assigning both male and female ranks, some differences in decision rules for male and female rank do exist. Refer to the Duke protocols for details on male rankings.

For the Amboseli baboon groups, we assume a linear dominance hierarchy. Each member in a group has a unique numerical rank that represents its position in the hierarchy relative to all other same-gender members of the group. Ranks are tabulated for groups by month, so month is the finest scale resolution available for rank. Female ranks are considered stable unless there is sufficient evidence to justify a change.

Assigning new ranks is both a science and an art. There is a clear mathematical component to it in that any rank reversals identify a potential change and having the smallest sum of reversals is theoretically the most correct matrix. (An individual can be beaten by any other baboon ranked above it, but if she is beaten by a lower-ranked member of her group, this may warrant a rank change.) Also, total number of wins and losses per dyadic interaction in that month often solidly dictates a rank change.

In situations where the data are scarce and the number of wins in a reversal does not clearly dictate a rank change, rank assignment becomes more of an art dictated by the following:

- Is the reversal offset that same month with a win by the other female?
- Are there baboons in between the individuals involved in the reversal?
- What happens to those "in between" individuals if you change a rank?
- How old is the female and where has she been ranked in the last years?
- Did she give birth that month? What is happening with her infant?
- Did she have any wounds or pathologies noted that month?
- Where are her mother and maternal sisters ranked?
- What interactions has she had in the last six months and the next six months?
- How large is the potential rank change (i.e. from rank 20 to 2 or from 20 to 19)?

When there is a need to look forward and the agonism data are not yet in Babase, review the monthly matrices that are sent from the field Team in Amboseli to look for pertinent trends.

Few set rules exist for assigning ranks. By considering interactions both in the immediate month and over longer periods of time and (to a lesser degree) considering an individual's family line and age, you begin to get a sense of what does or does not justify a change in rank position. The details are best sorted out during training or by referring to old matrices and considering the decisions made in the past.

Let's revisit the example above.

In our hypothetical example.
PEM>WEA 1:0
06A 3:2
06B 2:0

Let's say our matrix is for the month of January. PEM won over WEA 3 times between Jan and June, but PEM lost to WEA 2 times. The months the wins and losses occurred will dictate in which month to credit an individual with a rank change.

Jan 1:0	Jul 0:0
Feb 1:1	Aug 0:0
Mar 0:0	Sep 0:0
Apr 1:1	Oct 2:0
May 0:0	Nov 0:0
Jun 0:0	Dec 0:0

Both individuals seem fairly evenly matched until October. Based purely on the numbers, PEM should probably not be elevated in rank until October. However, you can ask the questions above and consider kinship and heath factors that may argue for elevating her rank as early as Jan or Feb.

However the matrix could instead be from June and the monthly data could look like this:

Jan 0:0	Jul 0:0
Feb 1:1	Aug 1:0
Mar 0:0	Sep 1:0
Apr 1:1	Oct 0:0
May 0:0	Nov 0:0
Jun 1:0	Dec 0:0

In this case PEM looked pretty evenly matched until June when PEM starts to win more often. Credit should be given in June, or even as early as April if Jeanne believes April marks the beginning of PEM's rise.

PEM (rank 6) could be moved to a rank 4 position, above WEA. WEA would then become rank 5. Individuals can move up in rank, but they can also move down. If you have a female who is consistently losing to lower ranked individuals during a particular year, instead of moving each individual up over the loser, consider moving the loser down below the winning individuals. This is particularly appropriate when the losing female is of high rank (for example, see WIRE in Weaver's group in 2006).

As you review each reversal each month, record your recommended rank changes for Jeanne with shorthand given for the win:loss values. Use the following language conventions:

No Credit – individual does not need to change rank
Give Credit – individual has enough wins to deserve a higher rank
Credit Already Given – an earlier rank change solved this reversal
Wait to credit – an individual should be credited with a rank change later this year

Write a brief explanation of your reasoning unless the win:loss values make the reason for your decision fairly obvious. Focus on questions above to justify your decision. Do not try to make snap judgments for reversals with limited data, or for Oct-December reversals when agonistic data for the next six months are unavailable. You can review the following year's matrices from the Amboseli Team for extra guidance, but it is best to flag anything questionable and revisit the decision in the next year with more data.

Once you have finished the recommended rank change for each month, return to *Ranker* and re-print the matrix for that month showing all the recommended rank changes from the previous months BUT NOT THE ONES YOU ARE RECOMMENDING THAT MONTH. For example, when you print the May matrix for Weaver's group, use Ranker

to show all your January-April rank recommendations so that Jeanne can see how your recommendations would play out in the month of May.

Remember, DO NOT save changes to Babase yet. You are just changing the matrix file that you print. Circle the sname and show the recommended rank change FOR THAT MONTH ONLY with an arrow to the new position. You will have to keep careful track of the rank changes you recommend. Remember your first set of matrices (v1) should always show the December rank order. It is from this matrix that the reversals for the entire year are identified.

HELPFUL HINT
Once you decide a rank change should occur, look at the cumulative matrix and go through the stickies for all the reversals in each month until the end of the year. If your decision has solved a problem or changed a reversal that exists in later months, mark the relevant reversal in the later month as "Credit Given". This will save you having to rethink the entire interaction every time you come to a new month with a new stickie, only to realize that you already solved that problem.

Updating Ranks
Once you have considered all months/groups, go over the matrices and decisions with Jeanne. After getting her approval to give or deny credit for each reversal, you are ready to return to the ranker program, run the matrix again, drag and drop individuals to their appropriate rank for that month, and save the rankings. Immediately after saving the all female ranks (ALF) for a group-month, you will also save the adult female ranks (ADF) for that same group-month by running the same matrix selecting only the ADF females in the early selection window.

MANIPULATING THE RANK ORDER (UPDATED 2009-06-18)

There may be times when you want to examine a monthly set of agonisms under the condition of a rank order from a specific time… even if that time was long ago. You do this by creating the matrix for your month of interest from the first set of options in Ranker. (See **Step 3: Create ALF Matrices for Each Group)**

Once you have the matrix in front of you, go to the menu option
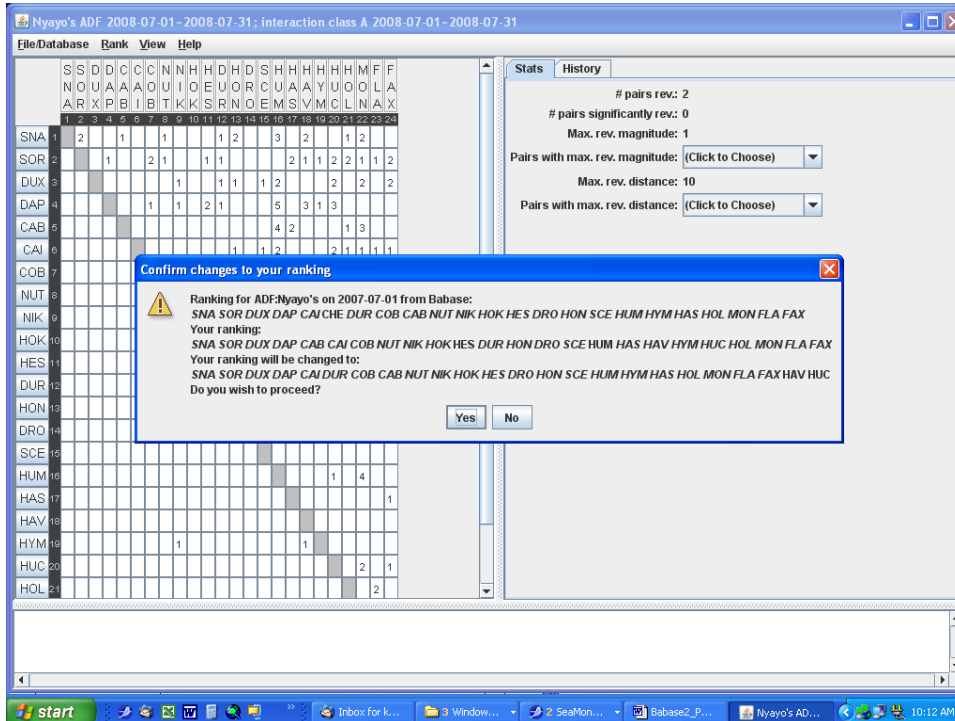Ranker > Load / incorporate ranking from Babase
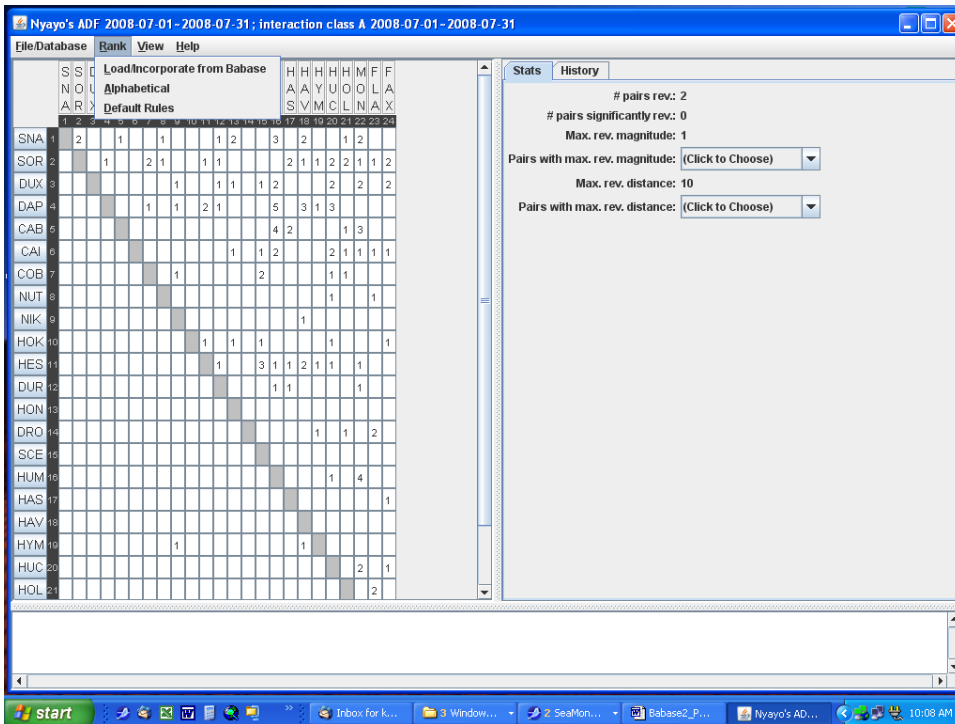
In the dialogue box select the options you want from the drop-down boxes. In the example below, I'm going to examine how the agonisms from July 2008 would look if I used the rank order from July 2007.
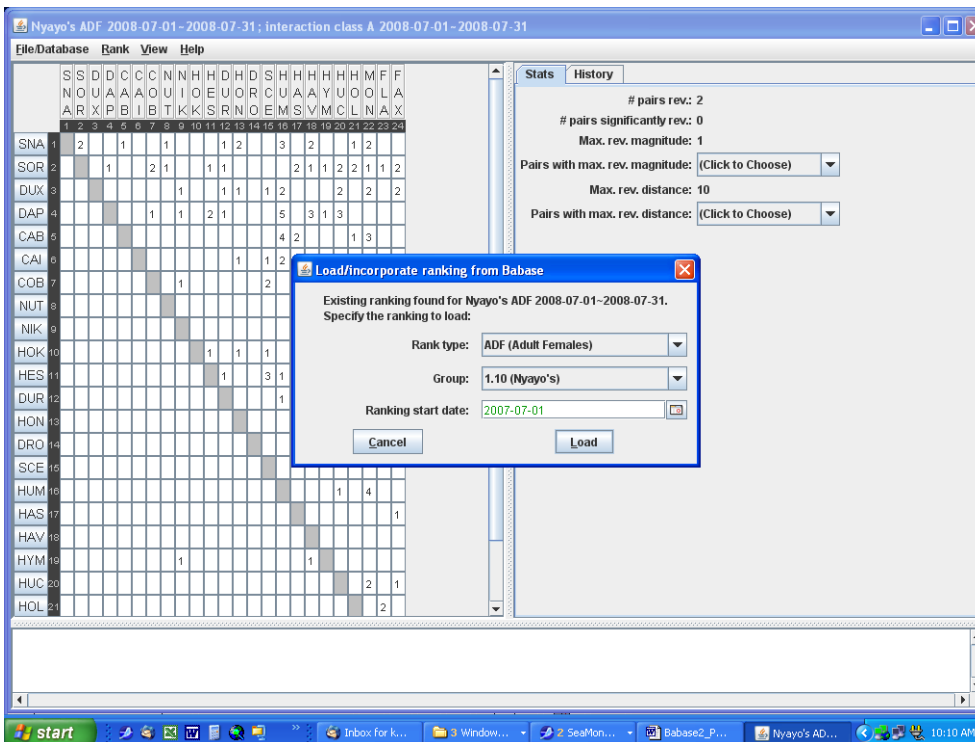
Click Load

You then get the following dialogue box to confirm the change.
Click Yes.

Ranker allows you to also do manual changes of an individual's rank in order to explore how wins/losses change within the matrix. This is useful in the second and third versions of the matrices as you begin manipulations of rank order. Drag and drop snames to move individuals up or down the ranks.

RANKING THE ADFs (UPDATED 2009-06-18)
Female ranks are entered and updated in Babase using the Ranker programs for all female ranks (ALF) and adult female ranks (ADF), respectively. The Ranker program displays decided agonism data for a particular group and time period in a win/loss matrix showing the number of times every individual beats another or is beaten in an agonistic interaction.

The menu option described above:
Ranker > Load / incorporate ranking from Babase
Is used to extract the ADF ranks from the ALF ones for each month

To do the ADF rankings, create a matrix for your month of interest using the ADF option:
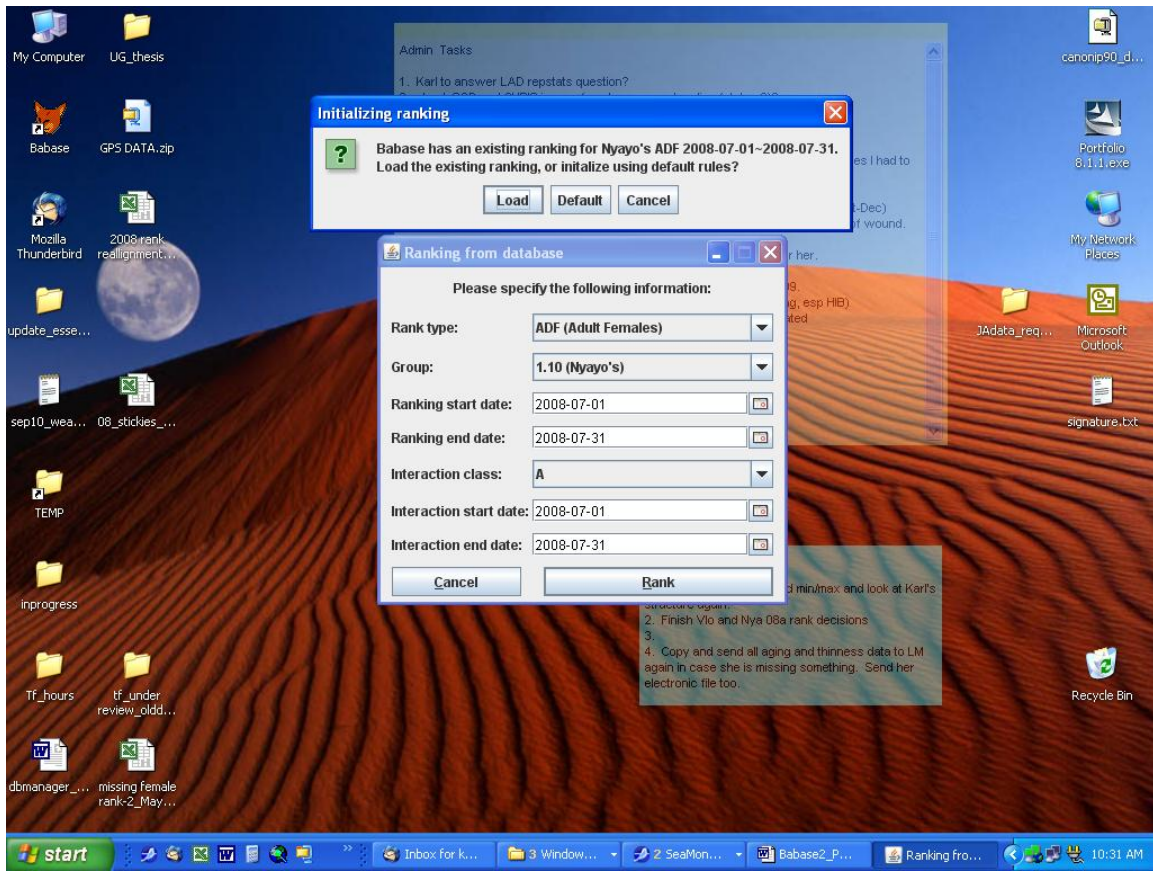
From the screen with the matrix use the menu option:

Ranker > Load / incorporate ranking from Babase

But choose the ALF as you ranktype. This tells Ranker you want to use the rank order from the ALF of the same month.
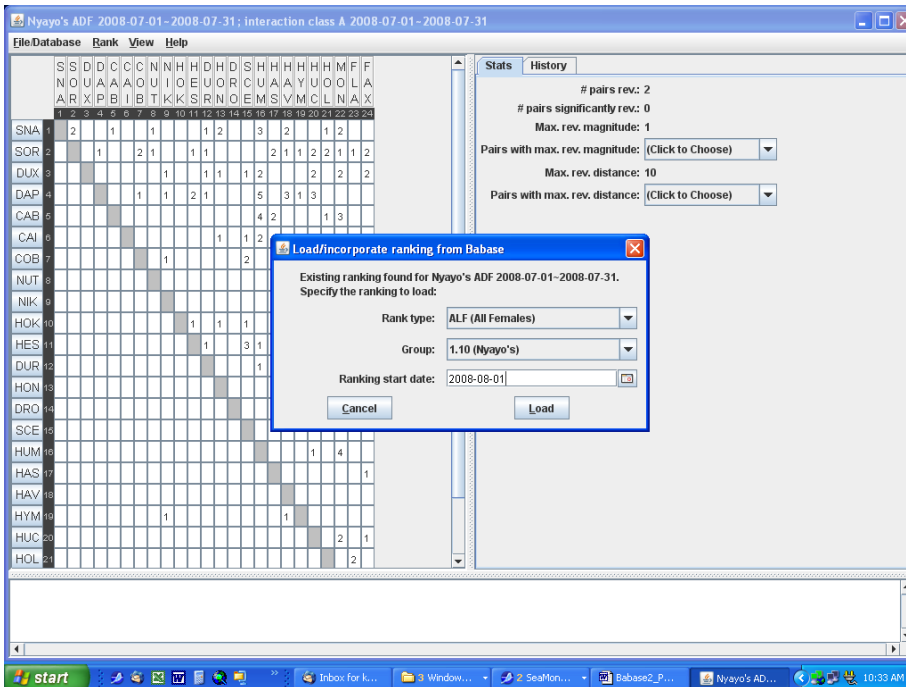
You then get a dialogue box that shows you the ALF ranks for the month you've chosen. Adults are shown in italics and juveniles are in block letters. The next few lines tells how your ranking will be adjusted to follow the order of the ALF ranks (minus any juveniles).
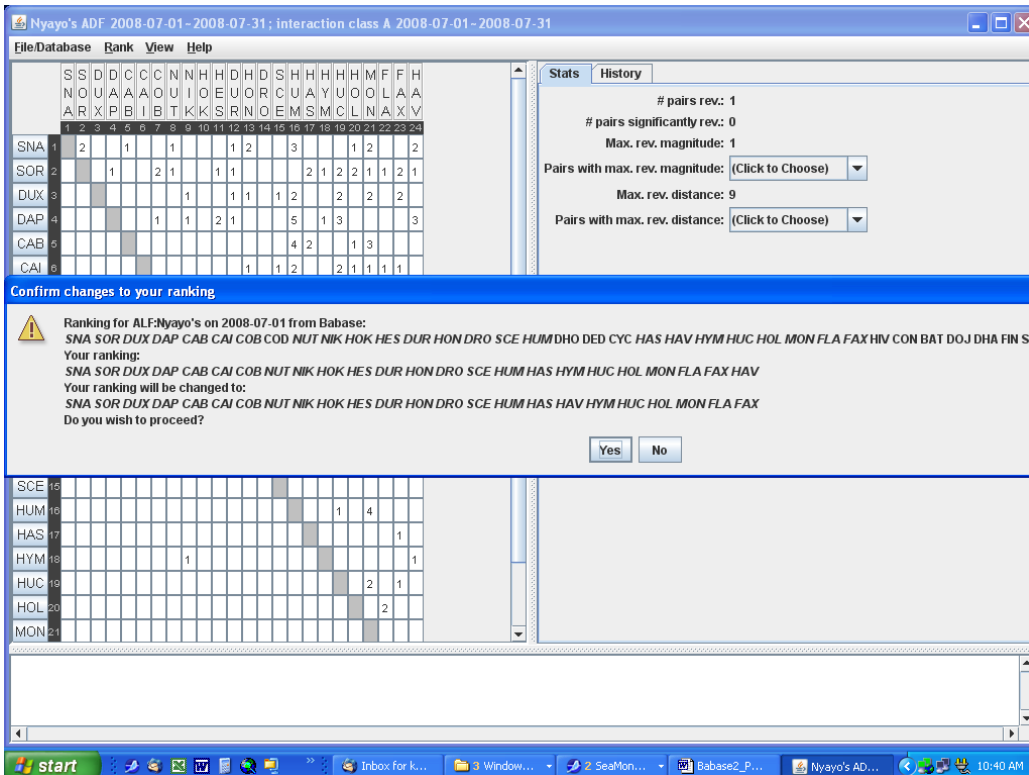
Note something interesting in the box above….HAV was at the end of the ADF rank matrix I first created because she matured in July 2008. Ranker automatically puts her at the bottom of the pack when she has her first maturedate. However, in the ALF ranks, she had already won several agonisms and moved up considerably. By telling ranker to seed the July 2008 ADF matrix using the already saved July 2008 ALF order, Ranker knows to put her in her proper place above HYM (but Ranker continues to leave out any immature females …those in block letters from the first line of the dialogue box).

Ranker simply extracts all the adult females out of the established all-female rank orders. The rank order of these females is preserved, though the specific rank number is adjusted to avoid skips. For example, suppose the top seven ranking females in a group are as follows (asterisks indicate adult females):

Sname	ALF Rank
AAA*	1
BBB*	2
CCC	3
DDD*	4
FFF	5
GGG*	6
HHH	7

The ADF ranks would be as follows:

Sname	ADF Rank
AAA	1
BBB	2
DDD	3
GGG	4

METEOROLOGICAL DATA

Handwritten Min/Max and Rainfall Data

Minimum and maximum temperatures and cumulative rainfall data are collected daily from a weather station at camp. This information is sent to Princeton in the monthly package and, having gone live on Babase in 2009, is updated every six months during the demography update. The data sheets have columns for date, time of reading, observer, minimum temperature, maximum temperature, precipitation, and notes, including days since last temperature and precipitation reading. These columns are described below along with conventions for data entry.

Date

Every calendar date from the first meteorological reading on 1 January 1971 to present is listed, regardless of whether or not data was recorded that day. If no measurements were recorded on a given date, "NR" for "no recording" is entered. Dates are in British format (day/month/year). All dates were included in an Excel sheet kept preveiously that was never integrated into Babase 1.0 in FoxPro; however, dates with no readings are not entered into Babase 2.0.

Time

Time refers to the time that the meteorological readings were taken. If the time field was left blank on the datasheet but measurements were read, "Unknown" is entered. If no measurements were recorded on a given date, "NR" for "no recording" is entered. Time units are in military format.

Observer

The observer column lists the individual(s) responsible for meteorological reading. Early in data collection, observer(s) were listed at the top of the monthly data sheets (therefore daily observers are not known). If the observer field was left blank on the datasheet but measurements were read, "Unknown" is entered. If no measurements were recorded on a given date, "NR" for "no recording" is entered. Observer column format includes two or three letter initials in alphabetical order separated by a comma.

Minimum Temperature (ºC)

Minimum temperature records the minimum temperature reading recorded on the corresponding date. If no minimum temperature was recorded on a given date, "NR" for "no recording" is entered. All temperature readings are in ºC (readings originally recorded in Fahrenheit have already been converted). For entry into Babase temperature is recorded in .5-degree increments only, as finer readings were deemed inaccurate.

Maximum Temperature (ºC)

Maximum temperature records the maximum temperature reading recorded on the corresponding date. If no maximum temperature was recorded on a given date, "NR" for "no recording" is entered. All temperature readings are in ºC (readings originally recorded in Fahrenheit have already been converted). For entry into Babase temperature is recorded in .5-degree increments only, as finer readings were deemed inaccurate.

Precipitation (mm)

This column is the precipitation reading on the corresponding date. If no precipitation was recorded on a given date, a "0.0" (zero) is entered. When precipitation was recorded as slight, very slight, trace or < 0.1, "0.0" is entered as the measurement reading – this was done for ease and consistency from the earlier dataset of 1971 through 1991. Also, if the reading was recorded as being between two measurements (e.g., "0.1 – 0.5"), then the lowest reading is entered (in this

case, "0.1"). All precipitation readings are in mm (readings originally recorded in inches have already been converted).

Days since Last Reading
The old Excel file had columns for Days since Last Temperature Reading and Days since Last Precipitation Reading. In Babase this has been altered to a column titled rgspan (rain gauge span), in which babase calculates and records the number of days, hours, minutes, and seconds since the last reading. An additional column, estrgspan, indicates whether or not the rgspan column was estimated (due, for example, to a missing time on the data sheet).

Note: For the most part, recordings are taken on a daily basis from a max/min thermometer and rain gauge at camp. However, skipped readings can be fairly common and it is critical to address the issue of missing readings (i.e. gaps in the daily recording schedule) when using this dataset. Generally, if readings were missed for one of more days, the first reading following the misses applies to all the missed days. For example, a precipitation of 20.0 mm after a four-day miss means that it rained 20.0 mm over four days. However, there is some variation depending on the observer that recorded the information. When using the data for analyses, just remember that some error is likely to occur when making assumptions about the data.

Uploading Min/Max Data to Babase
In July 2008, the data were uploaded to Babase 2.0 in the schema babase_pending for the first time. This data was fully integrated into babase in 2009. The min/max data will henceforth be uploaded to the view MIN_MAXS during the regular demography update or more frequently as necessary.

The types of errors that were encountered in prior uploads were:
  extra columns after notes which do not exist in the Babase 2.0 table (use the Tab button to check for these)
  "ASK" in time column not allowed: replace with NULL
  "unknown" in various columns not allowed: replace with NULL
  NR must be changed to "NULL"
  N/R must be changed to "NULL"
  change "unknown" in precip column to NULL
  delete extra rows at the end of the spreadsheet that created rows of NULLs
  In notes, putting a double space between sentences creates a higher row in Babase. Use one space only for better formatting when browsing the database.
  make sure all headers in smalls not caps and they match the Babase table headers exactly

Check for the above problems if you are getting error messages during new uploads. Also be aware that the new update worksheets may be formatted slightly differently. If the uploads are creating errors, also try playing with the formatting in Excel.

If for some reason you need to copy or move the table from your schema, you can use a command like:
create table Babase.min_maxs as select * from kfenn.min_maxs

Note that to copy tables in between databases (like Babase_test and Babase, you must dump the table and re-import it. The command above can only be used within schema of the same database.

METEOROLOGICAL DATA (PROOFED AND COMPLETE JUL 2008)

WeatherHawk Weather Station
The WeatherHawk Weather Station automatically takes hourly readings on a number of meteorological measurements. The Team in Kenya downloads this information into daily CSV files. These files, in turn, are emailed weekly to Princeton along with the other electronic datasets (i.e. point samples and GPS readings). The steps involved in managing this dataset used to include: inventory of data files, reviewing data content, prepping files for inclusion in the master spreadsheet, and maintaining a current master spreadsheet. There were WeatherHawk log sheets to fill in as you worked through these steps. The process is now simplified and the data uploaded to Babase_pending quarterly. Additionally, receipt of the data is confirmed on the Babase listserv on a weekly basis.

Step 1: Inventory of Data Files
As the zipped data files arrive each week in .csv format, download them to your desktop, upzip them to C:\Database Files\ALTMANN\Data from Amboseli in the appropriate monthly folder. The datafiles are generated automatically but it is important to review the data content to check the program settings. Make certain that units and date are British and time is in military format.

Originally the data were downloaded using third-party software (of which we have apparently used more than one version, including WeatherHawk PAKCOM, WeatherHawk XP, and WeatherHawk Virtual Weather Station Pro Edition v12.07; these shall henceforth be referred to as the older software) that downloaded each day's worth of weather data (0:00-23:00 hours) into its own .csv file. Later WeatherHawk provided us with in-house software, VisualWeatherHawk version 3.1. We began using VisualWeatherHawk in August 2010; however, since the WeatherHawk itself stores a great deal of data, we have data downloaded by the new software dating back to 23:00 hours on 22 Aug 2009. Consequently we have two versions of the data for approximately one year from August 2009 through August 2010. Whereas the old software downloaded each day's data into its own individual file, VisualWeatherHawk appends newly downloaded data to a single master file and the user can create output files for shorter periods so the Team is able to select a week's worth of data to send in one .csv output file. As always the files must be checked to be sure no dates or hours are missing. Since the files typically include only part of the last day, the hours most likely to be missing are those from the end of the prior week's file, which occurs if the current week's data begin on the following day instead of repeating the last date from the prior week. Sometimes whole days are also missing or the wrong file is emailed from the field. Be sure to check

the dates and immediately request the correct files from the Team if any days or hours are missing.

You will need to proof several of the raw data columns to try and catch any suspect data and alert the Team to any sensor malfunctions so they can be watched or corrected. This is best done once the files are in Excel format so do the following before proofing.

Step 2: Reformatting .csv Files into Excel– Protocol Used for Older Software
The .csv files that arrive must be 1) reformatted and 2) saved as .xls files before they can be merged to the master sheet. It is generally easiest to do a group of files together. Doing them at the end of each month rather than each week is recommended. This keeps the data relatively current, but saves a bit more time than doing them each week.

The .csv files have a header row for the date, but we need the date in the first column for each hourly record. There is a simple macro to help reformat these files. Follow these steps:

- Copy each .csv file you want to convert from
    C:\Database Files\ALTMANN\Data from Amboseli
  Into
    C:\Database Files\ALTMANN\New Babase Datasets\Meteorological\Daily Reformatted Excel WeatherHawk Data
- Open the file **macro2.xls** in this folder (enable the macros when prompted), then open the first .csv file you want to reformat.
- Place your cursor on the A2 cell and press Cntl+Shift+Q (this activates the macro) You should see the worksheet reformat and the date should now be filled in on every row in the A column.
- Save As a .xls file and close the data file (Cntl+W) without closing the window for the macro.
- Repeat with all the files for that month.
- Delete all .csv files from **Daily Reformatted Excel WeatherHawk** folder once you are sure you have a reformatted .xls version of them.

***Note: With the newer VisualWeatherHawk software, the files contain a week or more of data and the files do not require reformatting of header rows, thus they are not consolidated into monthly files. Instead the data are simply added to the master WeatherHawk file on a monthly basis. That file can be found here:
C:\Database Files\ALTMANN\New Babase Datasets\Meteorological\inprog_whawk\ABRP_weatherhawk_data_master_file.xlsx

The above file also includes the data downloaded with the older software but converted back to the units output by the WeatherHawk itself, which are the units reported by the newer VisualWeatherHawk data. Raw and corrected (to remove faulty rainfall and solar data) versions of data downloaded with the older software can be found in the same folder in the file called Corrected WeatherHawk collected with old software.xlsx.

Step 3: Proofing the raw data ****

The following table shows the WeatherHawk values and provides some guidelines for proofing the raw data. (**Measured versus calculated values?).**

OLD SOFTWARE COLUMNS AND UNITS

Wind Dir	°	No good way to know data, but wild swings may be suspect
Wind Spd	km/hr	Variable but generally winds may be stronger during the latter part of the day
Wind Gust	km/hr	Variable but generally winds may be stronger during the latter part of the day
Hum In	%	Not sure how this is calculated
Humidity	%	Higher at night than in the day except in wet months
Temp	°C	Look for hotter temps during the day and large highs / lows
Raw Barom	mm Hg	Watch for high or low values. Seems to be in the 660s -670s.
Tot Rain	mm	Cumulative annual rainfall – measured in 1mm increments
ET	mm	Evapotranspiration – according to WeatherHawk the calculation for this is set for a station set amidst a grass lawn and is thus not accurate in the Amboseli climate.
BattV	°	Battery voltage
Solar	W/m^2	Values should rise though the day and be 0 during night
Heat Ix In	°C	Not sure how this is calculated
Heat Index	°C	Not sure how this is calculated
DailyRain	mm	Cumulative rainfall for the day
HourRain	mm	Hourly rainfall
24HrRain	mm	Sum of rainfall from the trailing 24 hours
Deg Heat	°C	Not sure how this is calculated
Deg Cool	°C	Not sure how this is calculated
MonthRain	mm	Cumulative rain for the month
DegHeat Mo	°C	Not sure how this is calculated
DegCool Mo	°C	Not sure how this is calculated

NEW SOFTWARE (VisualWeatherHawk) COLUMNS AND UNITS

|		
RecNum		Reference number for the data row
BatVolt_V	volts	Battery voltage (hourly)
BatVolt_V_Min	volts	Minimum hourly battery voltage
AirTemp_C_Avg	°C	Average hourly air temperature
RH_Avg	%	Average hourly percent humidity
WindSpeed_ms_Avg	m/s	Average hourly wind speed
Solar_Avg	W/m^2	Average hourly solar radiation
ETo	mm	Evapotranspiration (great for a grass lawn)
AirTemp_C_Min	°C	Minimum hourly air temperature
AirTemp_C_TMn	time	Time of minimum hourly air temperature
AirTemp_C_Max	°C	Maximum hourly air temperature
AirTemp_C_TMx	time	Time of maximum hourly air temperature
WindSpeed_ms_WVc(1)	ms	Appears to be a repeat of Windspeed_ms_Avg
WindSpeed_ms_WVc(2)	°	Wind Direction
WindSpeed_ms_Max	ms	Maximum hourly wind speed
WindSpeed_ms_TMx	time	Time of maximum hourly air speed
Barometer_KPa	kPa	Atmospheric pressure

RainYearly_mm   mm   Cumulative annual rainfall (in 1mm increments)

You can do a quick and dirty visual check on most of the raw data columns by making some simple graphs.

To examine Wind Gust and Wind Speed, make a quick line graph. The daily wind variables from each column should track one another as in the following graph:
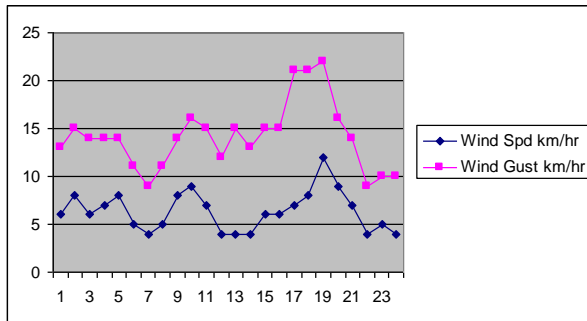

The Humidity and Temp variables should show predictable rises and declines and likely cross one another during the afternoon when the humidity has declined and the temp has increased (this will vary in the wet season). Typical temp range is in the 20s to 30s. Humidity is more variable daily and seasonally so pattern is probably more important than absolute value. Expect higher humidity in the wet months and lower in the dry months.
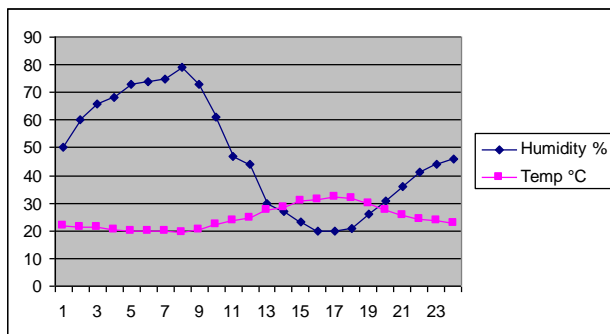
The barometer reading should show a predictable rise and fall pattern most days. The values typically fall near 670 mm Hg using the old software and near 89 kPa using the new VisualWeatherHawk software. Mean values have fallen over time and there seems to be a fairly large degree of variability among sensors. When the sensors were changed at the end of November 2008, a drop in barometric readings of approximately 2 mm Hg occurred. The patterns of change through the course of a day and over the course of the year (lower pressure during the rainiest months and higher during dry months) remains the same but the values dropped. WeatherHawk recommends that this sensor be changed

every four years. The sensor that originally came in the second WeatherHawk machine will be replaced in September 2013.
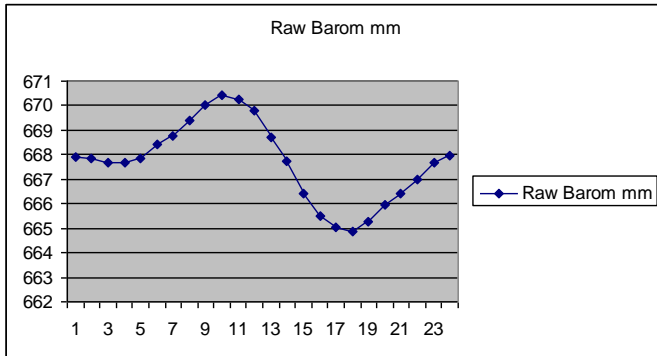
There should be a steady rise and fall of solar values. Daytime readings may be up around 1000 but it should go completely to 0 during nighttime hours. The curve will be fairly smooth on a sunny day but may be rather choppy, potentially with a much lower maximum value, on cloudy days.
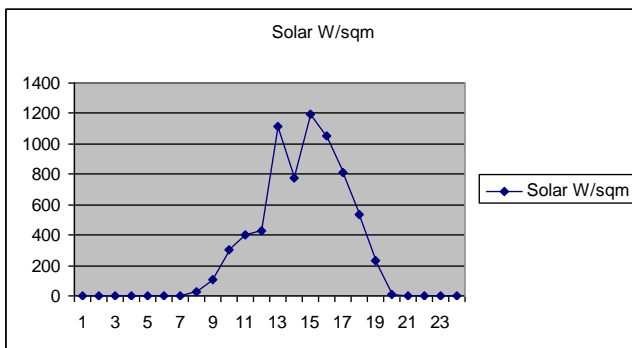
The old software had a variety of calculated rainfall columns, which resulted in many opportunities for mayhem in these columns. The new software only returns cumulative annual rainfall so as long as rainfall doesn't drop midyear, all should be well. For data collected using the old software, the following checks were performed. Visually check the **DailyRain** column and compare it to the other rain columns (graphs are generally unnecessary here as the rain levels are so often zero). All rain columns are cumulative over the course of a day, month, or year. **DailyRain** shows the cumulative rain received during each day with any new rain received during the hour added to the running total for the day. **24HrRain** shows the total amount of rain received in the previous 24 hours. Thus if there has been no rain for at least a day, when new rain occurs, the **24HrRain** column will initially mimic the **DailyRain** column, beginning one hour after the rain begins. And instead of restarting at the beginning of the new day, like the **DailyRain** column, **24HrRain** will continue to have readings until there has been no rain for the last 24 hours. **MonthRain** shows the cumulative amount of rain for each month and **Tot Rain** the cumulative amount for each year. These columns should increase each time it

rains and reset at the start of each new month or year, respectively. Occasionally one or more columns have malfunctioned for small or large periods of time. It is important to compare them to make sure the **DailyRain** column we use in Babase is working properly. The **HourRain** column does not appear to function correctly. We do not know why.

The rain bucket will tip and record a value only when it fills with 1 mm of water; consequently, there should be no decimal values in the rain columns. Also keep in mind that this method of measurement leads to frequent underestimation of precipitation in an arid climate like that in Amboseli.

If all the days and hours are present, and the values look reasonable, copy all the files for that month to their appropriate monthly folder within
C:\Database Files\ALTMANN\Data from Amboseli

Step 4: Consolidating the worksheets into a monthly spreadsheet – for older software only
> In the past we have paid for a license for Excel enhancer programs like DigDB. We have not maintained this license but you can download and run the macro for occasional use such as we have. Otherwise just use the Import tool in Excel. This takes about 10 minutes and is more under your direct control which has its benefits (you pay more attention to the steps of the process) but also its problems (you do more of the steps so there is more potential for human error). Or you can just copy and paste them into a monthly spreadsheet when you have each open for conversion from .csv to .xls and checking of data quality.

DIGDB instructions
> Copy all the PROOFED files for that month to the folder "monthly consolidated"
> Under the DigDB menu choose Table > Combine Files
> Click Add Files and go choose all the files you want moved into one workbook
> Click OK
> Make sure it imported all the worksheet you wanted
> Under the DigDB menu choose Table > Append Tables in this File.
> Choose All Sheets from the dropdown menu and click the add button.
> Click OK.
> Sort you new consolidated table by date and make sure all the days imported.
> Delete the first row that Excel automatically inputs.

Excel Import Tool Instructions (for Excel 2003)
- Copy all the .xls files you wish to merge into:
  C:\Database Files\ALTMANN\New Babase Datasets\Meteorological\to consolidate
- Open up the file for the first of the month and place your cursor a cell in the first row of the A column that is below the data.
- Go to Data > Import External Data > Import Data and browse to select the next .xls file you want to append to your file (i.e. the second of the month). Hit OK to

approve the selection and placement of files in the dialogue boxes. The next day's data should be imported below the prior day's data.
- Continue moving the cursor below the data and importing daily worksheets from each individual file into a single worksheet.
- Once all the files for that month are imported, freeze the top header and scroll down to each of the imported headers. Check all the headers against the top header to make sure the columns match. (Some time in 2006 the order and composition of the headers changed.)
- If all the columns match, delete the extra headers and rows leaving only the data.

Once the tables are in a single worksheet, then
- Check the **Time** column. If it looks wrong, reformat it using:
  Format > Cells > Custom > h:mm
- Add columns for **Month** and **Year** and populate them with the appropriate data.
- Insert columns as needed to make them match the master file.
- In a separate copy delete columns as needed to match the semiannual upload file.

Presently the WeatherHawk data reside in Babase_pending. There are two versions. One version contains all data collected using the old software. The second version contains all data collected using the new software and all additional data (for comparable columns) collected using the old software but converted to new software units. In addition to there being columns unique to each version of the software, there are some significant figure issues with conversion because of the differences in units and rounding performed by the old software. These issues will have to be resolved in some manner when WeatherHawk data are brought into real Babase and are as follows:
- Battery voltage was rounded to the nearest whole number by the older software
- Humidity was rounded to the nearest first decimal by the older software (but this is okay if we do the same in Babase – see the note below this list)
- Humidity was rounded to the nearest whole number by the older software
- Wind speed was recorded in different units (km/h relative to m/s) AND rounded to the nearest whole number by the older software, making it impossible to convert back to the original units with much accuracy (this one is a big problem!)
- Solar radiation was rounded to the nearest whole number by the older software
- Barometric pressure was recorded in different units and rounded to the nearest second decimal (but unlike with wind speed this does not create a sig fig problem)

NOTE: Dion Almond of WeatherHawk informed us upon query that the sensors are only accurate to the first decimal place, regardless of how many decimals the VisualWeatherHawk software provides.

Step 5: Appending the monthly worksheet to the master spreadsheet
For data collected using the older software, there are pre-made pivot tables and graphs in the Excel workbook that will automatically refresh with your new data if you make sure to add them above the bottom line of data. It was simplest just to insert the new data at the top. Then you could sort the data to move the new data to the bottom. Getting the

tables and graphs to update automatically would not work if you just pasted the new data at the bottom and was often problematic even then. At some point in time the below steps worked:

- Check the **Time** column. If it looks wrong, reformat it using:
  Format > Cells > Custom > h:mm
- Copy the function for the **Month** and **Year** column down into the blank spaces you just imported.
- On the pivot table worksheet, click the **!** icon (refresh) and see that your new month is included.

Although this seems like a pain, you can now update all the pivot tables and charts just by clicking the **!** button. ALL pivot tables and charts in the entire workbook will update with the new month's data and that saves you A LOT of time.

NOTE: If you end up with a new column in the pivot tables labeled "blank" go back into your worksheet and try deleting some of the rows at the bottom of the page (even if they appear empty). Then **!** (refresh) the data. The blank column should disappear. If you try this and it doesn't work, you have an empty row somewhere else in the dataset.

UPDATE: These tables are not in use with the WeatherHawk data collected using the VisualWeatherHawk software. In addition to issues with auto refreshment of the tables and the fact that many of the tables were running out of room physically within the worksheets, the differences in columns and units between the two versions of the software left the tables incompatible with data received in the new format. Should the need arise, new tables can be created using the columns and units from the new software.

Step 6: Uploading new data to the whawk table in Babase
Once the new data are entered into the spreadsheet there are two different ways the online database can be updated. Whichever you choose, first test it in your own schema or someplace like Babase_test before running it in the true Babase schema.

Strangely there is a blank table called whawks in Babase. I do not know why this is, as WeatherHawk has never been properly designed (though perhaps it was briefly started upon by T. before min_maxs was finished) and now it will need to be completely different due to the improved software.

WeatherHawk data are currently entered in Babase_pending. The data collected using the old software is in the table called weatherhawks_old and the data collected using the new software, as well as data converted from the output of the old software, is in the table called weatherhawks. The weatherhawks table is currently updated in quarterly installments.

Below are some tips for updating:

1) Empty the entire whawk table in Babase database, Babase_pending schema and reload it with the entire whawk table (going back to 2004). First be sure to:
   - Change all the headers in your Excel file to exactly match those of the Babase table (you can just copy the header row from C:\Documents and Settings\altmann lab\Desktop\inprogress\whawk_test_upload
   - Fill in the final column (ev) with NULL values if they are all blank (the upload program doesn't always know there is a final column unless the NULL values are there and it complains about too few columns)
   - Export the new updated Excel sheet to a .txt file.
   - Use the IMPORT button in PPA to import the entire file again (check the empty string and NULL boxes before importing)

2) Upload only the most recent months' data from an Excel file using the UPLOAD program or the IMPORT button. UPLOAD IS PREFERRED BECAUSE IT GIVES YOU A ROW COUNT AND RECORDS THE TRANSACTION. First be sure to:
   - Change all the headers in your Excel file to exactly match those of the Babase table (you can just copy the header row from C:\Documents and Settings\altmann lab\Desktop\inprogress\whawk_test_upload).
   - Export the new updated Excel sheet to a .txt file.
   - In UPLOAD, select the correct database, type in the table name, check the NULL box, and browse for your .txt file.
   - In IMPORT select TABBED for format, and check both the NULL and Empty string/field boxes before attempting the import.

NOTE: No matter what method you choose for the update, you must delete the month and year columns because they do not exist in the whawk data table in Babase and you will get an error if you try to upload them. We maintain them in the Excel sheet so we can make the pivot tables. This does not apply to data collected using the VisualWeatherHawk software, which provides the date and time in date/timestamp form.

VACUUM afterwards! In Babase_pending you have to do it as an admin!

MALE AGE ESTIMATES

Male age estimates, both age estimates for new immigrant males and annual age estimates on resident males (usually conducted in Oct), are stored in the following Excel file: C:\Database Files\ALTMANN\New Babase Datasets\Aging\Male Age Estimates. These tables are updated as data comes in and the data are periodically supplied for research questions. There are separate tabs for estimates for new immigrant males and for each year's annual age assessment.

WOUNDS AND PATHOLOGIES

Date: 26 September 2005 (updated 3 July 2006 and 24 Sep 2009)

Summary: The Wounds and Pathologies dataset records data documenting field observations of wounds and pathologies as well as notes on the animal's subsequent condition or recovery related to the wound/pathology. For entry, data from wounds/pathology field notes are separated into 4 related tables:

1. **Index Table** – Provides basic summary information about the wound/pathology. One row corresponds to every wound/pathology datasheet.

2. **Wound/pathology Code Table** – Lists specific wounds and pathologies associated with a row in the index table; provides additional specifics. One row corresponds to every wound and/or pathology code associated with a row in the index table.

3. **Body Parts Table** – Lists the specific body part(s) affected by the wound/pathology. One row corresponds to every body part associated with a row in the index table.

4. **Healing Rate Table** – Documents the healing status and dates for each wound/pathology listed in the index table. One row corresponds to every date of a follow-up comment.

Index Table – Columns:

WID
Every wound/pathology is given a code specific to a particular wound/pathology. The original observation and subsequent follow-up entries should all reference the same code. WIDs are generated manually and large gaps exist in the numbering sequence.

Date
The date of the initial observation. Dates are in British format (day/month/year).

Time
Time refers to the time that the initial observation was made. Time is recorded in military format.

Observer
The observer column lists the initials of the individual(s) responsible for the observation. Separate multiple observers by slashes.

Sname
Sname is the sname of the baboon with the wound/pathology.

Group
Group is the group ID of the group that the individual was in at the time of noticing the wound/pathology.

Wound/Pathology
The wound/pathology column refers to whether the wound/pathology row refers to a wound, a pathology, or both. Possible code values are:

Wound/Pathology Index	
A	Wound
B	Pathology
C	Wound and Pathology

Comments
The comments field is for any comments describing the initial observation of a wound/pathology.

Healing Rate Table – Columns:

WID
This column links the healing rate information back to the Index Table.

Date
The date(s) of the follow-up observation(s). Dates are in British format (day/month/year).

Healing Status
The codes for healing status reflect how the wound/pathology has healed on each follow-up observation date. Possible code values are:

Healing Status Codes	
1	Not healed
2	Partially healed
3	Healed
4	Terminal
5	Animal missing

NOTE: If a WID refers to more than one wound and they heal at different rates, the healing status codes should reflect the slowest wound(s) to heal.

Body Parts Table – Columns:

WID
This column links the body part information back to the Index Table.

Body Part
Body part locates where the wound or pathology is on the baboon. Not all wounds/pathologies are associated with a body part. Possible code values (in reference to baboon drawing) are as follows:

Body Part Code	
0	Head (unspecified)
1	Top of head
2	Eye region (includes brow)
3	Muzzle
4	Lower jaw, mouth, throat
5	Cheek
6	Ear
7	Back of head, sides of neck
10	Arm (unspecified)
11	Shoulder, armpit
12	Upper arm, elbow
13	Forearm
14	Hand, wrist
20	Trunk (unspecified)
21	Ventrum, chest, between legs
22	Flank
23	Upper or mid back
24	Lower back
25	Sacral region
26	Hindquarters, PCS
27	Sex skins
28	Genitals
30	Leg (unspecified)
31	Thigh, knee
32	Lower leg
33	Foot, ankle
40	Tail (unspecified)
41	Proximal Tail
42	Tail hook
43	Distal Tail
44	Over entire body

Body Part Side
Body part side indicates whether the body part affected was on the left (L), right (R), or center (C) of the animal.

Wound/Pathology Table – Columns:

WID
This column links the body part information back to the Index Table.

Wounds/Pathology Code
The wounds/pathology code refers to the specific code(s) listed on the datasheet.
Possible code values are:

Wounds/Pathologies Codes		
Wound	1	Linear cut or slash
	2	Puncture
	3	Scrape, amorphous wound (includes tick wounds)
	4	Bruise, swelling
	5	Multiple small cuts (all < 1 cm)
	6	Other wound (e.g., dislocations)
	16	Unknown/ indiscernible wound type
	17	Scalping/ large, open wound on head exposing the skull
	18	Broken bone
Pathology	7	Limp, no wound visible
	8	Respiration problems, coughing, sneezing
	9	Digestive problems, vomiting, diarrhea
	10	Malaise, weakness, stiffness in absence of wound
	11	Thinning fur
	12	Nosebleed
	13	Discharge, sores, rashes
	14	Other pathology (e.g., low body weight, blindness)
	15	White fur

Maximum Dimension (cm)
The maximum dimension of the wound (in centimeters).

Impairs Locomotion
A yes or no field for whether or not the wound/pathology impairs the animal's
locomotion. **NOTE:** Entry of a "yes" for impairs locomotion is inferred if notes field on
datasheets explicitly describes a locomotion problem and/or if "limp, no wound visible"
is checked on the wounds/pathologies list.

Signs of Infection
A yes or no field for whether or not signs of infection (oozing, redness, or stiffness) were
observed. **NOTE:** Entry of a "yes" for signs of infection is inferred if notes field on
datasheets explicitly describes an infection associated with a wound. Note also that this
field is for signs of infection at any point during the monitoring of a wound – it is not
limited to wound status on first observation.

Data Entry Tips and Tricks

Overview. Entering the Wounds and Pathology data may sound complicated, but it really isn't bad once you get the hang of it. The most important thing to remember is keeping track of adding all the information to all the possible tables – if you stop for a break, try to avoid leaving entries half-finished. And save the spreadsheet often – nothing is more frustrating than having to enter the same information twice!

Before you even open the spreadsheet and consider starting data entry, take a look at one of the Wounds and Pathology data sheets. A sample one is scanned in below. Familiarize yourself with the type of information collected and what constitutes a new form. You should notice that each datasheet is specific to one animal, but not necessarily a single wound or pathology. For example, in the sample below, on 19 August 2004, Orion was observed with both a wound (scrape, amorphous wound) and a pathology (limp, no wound visible). All follow-ups after the initial description of a wound/pathology are written towards the bottom of the data sheet. Keep in mind that the number of follow-ups and descriptions associated with each follow-up varies somewhat. In the example below, there were two follow-ups to the wound and pathology originally recorded for Orion on 19 August 2004 – one follow-up on 28 August 2004 and another on 28 September 2004. In theory, follow-ups should continue until the wound/pathology is healed or the animal dies. In practice this does not always occur.

Data Entry. In the paragraphs and tables which follow, I'll refer to the two sample wound/pathology sheets for Hasty and Nafuu (scanned below) and describe how the data get entered into the various spreadsheets. Before reading on, be sure to look at the written description of the Wounds and Pathologies datasheet for specifics on spreadsheets, columns, codes, etc.

Scanned sample wounds/pathologies sheets:

Index Table:

WID	Date	Time	Observer	Sname	Group	Wound/Pathology	Comments
1	20-Mar-03	15:03	JKW	HAS	1.10	B	Hasty limping slightly with her right arm
2	7-Feb-03	13:00	JKW	NAF	1.10	C	Nafuu is limping badly from a thigh wound that he probably got from fighting with other males. The wound is also bleeding and it is on the left leg thigh.

- Be certain that you are assigning a UNIQUE WID – otherwise we won't be able to accurately reference different data tables. If you make a mistake and need to delete a WID, that's ok but do not worry about filling in gaps. It is ok if WIDs skip a number.
- Remember that the date is in British format.
- Enter the time in military format with a colon.

Healing Rate Table:

WID	Date	Healing Status
1	29-Mar-03	3
2	27-Feb-03	1
2	29-Mar-03	1

- Again, remember that the date is in British format.
- On this and all subsequent tables, check and recheck that you are keeping the WID consistent!
- If there were never any follow-ups made on a wound/pathology, you do not need to enter any rows in this table.

Body Parts Table:

WID	Body Part	Body Part Side
1	10	R
2	31	L

- If no body part applies to the injury (for example, general weakness), then don't add any data to this table.
- If more than one body part applies to an injury, put each body part in a separate row
- If a wound occurs on both the left and right side of the body, enter two separate rows for "R" and "L" respectively

Wounds and Pathologies Table:

WID	Wound or Pathology Code	Maximum Dimension (cm)	Impairs Locomotion	Sign of Infection
1	7			
2	1	4	Y	
2	7		Y	

- It is not unusual for a single datasheet to reference several wounds/pathologies. That's no problem – simply create a row for each unique wound or pathology code.
  - In these cases, it may be difficult to determine which wound/pathology the listed wound dimensions, impairs locomotion, and sign of infection notes apply to. The written comments often provide key clues in determining this.

Additional Notes. Below is text written by L. Okpala on some additional data-entry rules.

- When body parts fell off/out (*e.g.,* canine, eye, tail), the entry was marked as healed and a note was made of it.

- A wound/pathology deemed as permanent was said to be "healed".

- If wound had been reopened and a separate record sheet was NOT made for it then the wound was marked as partially healed (2) and a note was made that another new injury existed.

- If a separate data sheet was made for the reopened wound, then the wound was treated as a new wound and an entirely separate entry was made.

- If a wound was not classified as reopened but two separate data sheets were made at the same time but for different wounds, both the data sheets were combined and a note made in the data book.

- If infection and impairment of locomotion could be inferred from the data sheets it was checked off and marked on the data sheet.

- If limping was checked on data sheet but comments inferred that limping was caused by wound, then the checked off "limping, no wound visible" was disregarded and "impairs locomotion" was checked off.

- Assignment of partially healed, healed and not healed was inferred from the notes. If "healed" put down twice for separate dates then each date got a separate entry.

- If limping from wound is noted after initial injury, then check "yes" in the impairs locomotion category.

- Sometimes data sheets will say "limp or stiffness from wound" rather than "limp, no wound visible". Take note of this and if "limp from wound" is checked off, then disregard this and check off "impairs locomotion" in addition to noting the original wound.

- All instances of fur abnormalities are included under the "thinning fur" (11) category. This includes, but is not limited to, notes citing things such as "a rough coat," "a bare patch of fur," and/or a region of the body missing fur.

White Monkey Syndrome

Any reference to a stiff walk related to white monkey syndrome will be assigned a pathology code of 10 (malaise, weakness, stiffness in absence of wound). If a baboon has both a stiff walk and white fur, both codes 10 and 15 should be marked.

INTERGROUP ENCOUNTERS

Date: 2 November 2009 by Catherine Markham

Note: Intergroup Encounter data currently reside in babase_pending and are updated quarterly.

Summary: The Intergroup Encounters dataset records data documenting field observations of group-level interactions.

Entry Notes
1. Only group level data involving two or more groups entered (i.e. excluded data on single groups and data on individual or males, unless identified as subgroups)
2. If time range was given, entered the start time
3. Checked for notes in Other Groups section *and* Demog Notes

Entry Tips
- "Following" typically considered an agonistic interaction with the dominant group being the pursuer, unless additional notes indicate otherwise
- Simple "approaches" are not considered agonistic interactions (unless additional notes – such as an approach followed by a displacement – make agonism clear)
- The activity of simply "watching" is not considered an agonistic interaction
- Care must be taken to interpret displacements and changes in direction (both sign of agonistic interactions) from baboon groups simply moving apart at their own will
- Field notes occasionally state that "no interaction seen" despite otherwise clear records of a group-level interaction; this is interpreted as no *direct contact* and/or *intermingling of groups* observed
    - Stating "no interaction seen" does not necessarily mean that a group-level interaction did not occur
- If field notes indicate that interaction was clearly agonistic but winner is not certain, mark "Y" for Decided Agonistic Interaction but "Uncertain" for both dominant and subordinant group ids
- It is not necessarily considered an agonistic group-level interaction if only a single male's behavior was noted as being aggressive (or submissive), although the record may likely be an indication that groups were close to one other
    - Group proximity *not* assumed when reference is to an adult male (since the adult could have been dispersed)
    - Group proximity is assumed when reference is to pre-dispersal males or females (since the rest of their group is likely nearby)
    - In certain circumstances, a subset of individuals may be responsible for a group-level interaction (*e.g.*, some males may be individually identified as running towards another group and causing that group as a whole to move away)
- Interactions involving non-study groups may contain estimates of groups size; include this data in the "Notes" field as well as qualifying notes of census condition
- Rarely, a single intergroup encounter note will include a dominance reversal (loss of one group followed quickly by the win of that group) – in these cases, the "winner" is considered the last group to be dominant
    - If groups eventually settle into apparently neutral or peaceful interaction in the aftermath of an agonistic encounter, the encounter is still coded as agonistic
- Note if observer's presence seems to have influenced the interaction

- Marking a row as "no" for whether or not the event was an agonistic interaction does *not* necessarily imply that the interaction was friendly – *i.e.,* the groups may have been neutral towards one another
  - What it does imply, however, is that the groups were within visual and/or vocal proximity to one another; groups within 500m of one another may be considered to be close
- Enter interactions involving more than 2 groups as all possible group-pair combinations
  - Enter the total number of groups involved in the column with this name
  - Note that in multi-party group interactions, some dyads may have interacted agonistically while others did not
- For time periods when it was uncertain whether certain non-study groups had fissioned (*e.g.* Sinya), only "Sinya group" was marked appropriate group column despite the fact that the interaction description may specify "*a* Sinya group"
- Field for Observation Certainty marked as "uncertain" when observers indicate they are *not* positive of the behaviors observed (*e.g.*, a group *seemed* to be pushing another group, a group *may have* changed direction) or if it is possible that observer's presence or presence of Maassai specifically influenced the outcome of an agonistic interaction (because one group acted subordinate due to poor habituation)
- When same interaction noted in several notebooks, field notes combined to give the most detailed account of the interaction possible
  - Make mention of the fact that note appears in multiple notebooks
  - Only enter the interaction once (do not repeat the same interaction for each notebook it appears in)
- Just looking nervous or appearing to be disturbed in and of itself does not constitute submissive behaviors
- Use country codes (KE for Kenya and TZ for Tanzania)
- Only enter records where it is clear that, at a minimum, two or more groups were in close proximity to one another
  - Several times, the team will record when they spot another group – this does not necessarily infer that this group was near the group they were observing that day (could have been found on the team's way to or from the focal group)

Entry Conventions
- Use "full name" for group (e.g., Dotty's group, Linda's group, Hook's group, Ositeti group)
  - Note that group can be identified as "Unknown" if observers were not able to make a positive group id
- Use snames for individuals
- Use the following full names and spellings for locations
  - Sinya Hill
  - Nado Soito Hill
  - Nairabala Hill
  - Naripi area
  - Illmerishari Hill
  - Napapong Hill
  - Ndoroboni Hill
- "Rain pool" is two words (not "rainpool") and "waterhole" is one word (not "water hole")
- Do not insert a space between distance measure and units (*e.g.,* "300m" *not* "300 m")
- Use brackets ("[text]") when adding a note or clarification upon data entry that was *not* in the original field note

- Subgroups only referenced in Notes field (entire group ID listed in appropriate group columns)
- The following are considered synonymous with "uncertain":
  - Unidentified
  - Unknown

Potential Improvements/Additions
- Though captured in the text of the Notes field, I did not pull out info on (1) end times of the interaction, (2) group size at time of interaction, and (3) minimum distance separating groups, when available, into separate fields
- It may also be interesting and relevant to mark records where interaction occurred with reference to a discrete ecological resource (waterhole or sleeping grove)

Considerations before Babase Upload
- Convert group names to group ID codes
  - Consider how to handle groups during fissioning events as well as groups such as Sinya group during the time period where the team was unsure whether or not the group had fissioned or other cases where the group ID applies to more than one group
    - Reference to the text in the Notes field will be very helpful in sorting out these specifics

HYBRIDITY

Hybridity used to be handled at Duke but ca. 2011 was passed to Princeton. Data entry was caught up and older data consolidated at Princeton. The complete raw hybrid scoring dataset was added to babase_pending in early 2014. This dataset can be found in the table hybridity_rawmorpho and is updated quarterly.

Raw hybridity scores can be used to calculate the morphological scores found in the babase_pending table hybrid_morpho. The only scores presently in that table were calculated by Lacey and appear to draw on raw data through July 2006. To date no additional scores have been calculated. I expect this will be done upon request. The current hybrid_morpho table does not have a column for iterations. This column should be added or subsequent iterations should be included in a separate table. The protocol for calculating these scores is included below.

Protocol for Assigning Final Morphological Hybrid Scores
Created: 23 January 2008
L. Maryott

Step 1: The first step once the raw data scores from the field are received is entering them in a systematic way into the spreadsheet. The image below demonstrates the proper format for the data sheet. The headers should be group, date, observer, sex, dob, age, category, sname, color, hair length, body shape, head shape, tail length, tail bend, muzzle skin. Juveniles should only have the first 5 of the 7 criteria scored, while adults should have scores for tail bend and muzzle skin as well.

Microsoft Excel - HYBRIDS_AUG_TO_DEC_06.xls

File Edit View Insert Format Tools Data Window Help

D100

	group	date	observer	sex	dob	age	category	sname	color	hair length	body shap	head shap	tail length	tail bend	muzzle sk
2	1.1	01/08/2006	SNS				JUV	HAD	0	0	0	0	0		
3	1.1	01/09/2006	SNS				JUV	CDO	0	0	0.5	0	0		
4	1.1	01/09/2006	SNS				ADULT	SEI	0.5	0	0	0	0	0	0
5	1.1	01/09/2006	SNS				ADULT	BUC	0.5	0	0.5	0	0	0	0
6	1.1	01/09/2006	SNS				ADULT	DUR	0	0	0	0	0	0	0
7	1.1	01/09/2006	SNS				JUV	COT	0	0	0	0	0		
8	1.1	01/09/2006	SNS				JUV	SEE	0	0	0	0	0		
9	1.1	01/10/2006	SNS				ADULT	HAS	0	0	0.5	0	0	0	0
10	1.1	01/10/2006	SNS				JUV	SAS	0	0	0	0	0		
11	1.1	01/10/2006	SNS				ADULT	SNA	0	0	0	0	0	0	0
12	1.1	01/11/2006	SNS				JUV	SEB	0.5	0	0	0	0		
13	1.1	01/11/2006	SNS				JUV	DAP	0.5	0.5	0.5	0	0		
14	1.1	01/11/2006	SNS				JUV	FAM	0	0	0	0	0		
15	1.1	01/11/2006	SNS				JUV	DHO	0	0	0	0	0		

Step 2: Once the scores are entered, a mean per observer, per date should be calculated. This results in one mean per row. These scores should be the average of either 5 or 7 scores, depending on whether the score is an adult or juvenile score.

> Note: If a Juvenile is scored on all 7 criteria for some reason, ensure that the team still considers the individual a juvenile, and then delete the last two criteria, as they aren't able to be accurately measured before adulthood.

Step 3: The means created in the last step (one mean per row) will result in between one and four mean scores per time period. I.e., some individuals are not scored by all three field team members, but for some, SCA scores the animals in addition to JKW, SNS, and RSM. These means (one mean per row) should then be averaged across individuals within a date to create a mean value across observers for each date on which the animal was scored. Sometimes the dates are not exactly the same for each observer; all scores collected within one month of each other on the same animal (for juveniles) or six months of each other (for adults) should be considered as collected on the same date. This results in a "mean of means" for each animal at each age or scoring event. These means of means will be the average of between 2 and 4 scores; . If there is only 1 score for a particular time period then just use that score, it will be a simple mean rather than a mean of means but will be treated the same way as the means of means in subsequent steps. [Is this on the same spreadsheet at step 2 or a new spreadsheet? Describe columns]

Step 4: The next step is to average all of these means of means for an individual during its juvenile life or its adult life to create a juvenile hybrid score and/or an adult hybrid score. This average could have any number of components, from 2 means of means, to 5 or 6 for natal males' juvenile scores. [What does this table look like? What are its columns?]

> Note: In the Tung et al 2008 manuscript, these are the scores that were used as final morphological hybrid scores. Every individual was assigned its adult hybrid score as the final score, unless the adult score was impossible to ascertain in

which case their juvenile score was used. These are the scores that should probably be given to inquiring researchers, although it is important to know their question before making this decision. Also, if distributing these scores, any juvenile scores which must be used should be flagged as such.

Step 5: The final step of hybrid scores, if desired is creating an overall morphological hybrid score by averaging the adult and juvenile scores. There are potentially two ways to go about this. There is the option of just average the final adult average score and the final juvenile average score. Since a lot of averages are contributing to these numbers, a potentially better, and more accurate option may be to instead average all of the 'age means'. In this case, anywhere from 2-25 scores would be averaged together. This may produce a more accurate final life score. [Check Marie's manuscript] [I think that we may not want to add this step. This could be up to individual researchers to decide to do this]

NEONATAL ASSESSMENT SHEETS

Date: 20 May 2014

These data were entered into Excel in 2011 by Kerri Smith and can now be found in the babase_pending table neonatals, which is updated quarterly.

For every infant seen alive in a study group on at least two census days you should receive two neonatal assessment sheets. There are two types, those for the First Observation and those for the Second Observation. The main different between the two is that the first asks more questions, those having to do with when the mother was last seen without the infant, first seen with the infant, the appearance of the infant and umbilical cord and whether the mother was bloody on various areas of her body. These questions and sometimes the Remarks are useful in estimating the age of the infant when first seen and thus in assigning a birth estimate. The first and second neonatal assessment sheets are entered into the same table and items not appearing on the second neonatal sheet should be left blank, as should items that are not filled in on either assessment sheet, such as information about the umbilicus, which is only filled in when an umbilicus is present.

DIGITAL PHOTOGRAPHS
Created by Catherine Markham, August 2006
Modified by Oriana Poindexter, October 2008

Note: Some photos (certainly not all those contained in the Princeton database – perhaps they are only recent photos?) are now archived in Photo Organizer, accessible on the wiki.

Adding New Images to the Photo File
Most digital photos arrive from Kenya on a CD included in the monthly data package. Review each CD and transfer new digital pictures to the appropriate subfolder in C:\Database Files\Altmann\ABRP Photos. Separate subfolders exist for every month/year and follow the naming convention below:

Y Y Y Y _ M M

YYYY = Four digits denoting the year. Use 1999 for 1999, 2000 for 2000, 2001 for 2001, etc.

MM = Two digits denoting the month. Always use a leading zero for months with MM less than 10. Use 01 for January, 02 for February, 03 for March, etc.

Cataloging Photos with Portfolio

With the use of a program called Portfolio, digital photos from the Amboseli Baboon Research Project can be easily reviewed and cataloged. In addition, Portfolio's search and find queries allow photos to be quickly located for inclusion in various presentations, posting on the website, etc. Refer to the Portfolio User's Guide for specific instructions on program features and use.

After all new pictures have been copied from the CD into the appropriate subfolder(s), open the Portfolio program and begin cataloging each image. Portfolio will automatically record imbedded photo data (such as date picture taken, camera model and type, aperture, etc.). In addition, we manually catalog the following fields:

Keywords: The Keywords field provides basic cataloging information about the photo's subject. Every photo should have at least one keyword and more than one keyword may be associated with a single photo. Acceptable keywords appear in a drop-down list and include:

Baboon – Photos of baboons.

Camp – Photos of life at camp in Amboseli.

Fieldwork – Photos of people doing fieldwork.

Labwork – Photos of people doing lab work or office work.

Habitat – Photos of habitat. Often these are landscape shots (do not include zoomed in pictures of specific plants).

People – Photos of people.

Plants – Photos of plants.

Wildlife – Photos of wildlife other than baboons (e.g. elephants, wildebeests).

People Names: The People Names field identifies individuals in the photo. The names of individuals with long-term involvement in the project are listed in a drop-down menu. The names of individuals rarely photographed or involved in the project temporarily (such as Jeanne and Susan's students) may be added to this list as needed. These names

will not be included in the drop-down list but will be queriable. When adding new names, be sure to follow the naming format of "Last name, First name" already established. More than one name may be associated with a single photo.

Baboon Names: The Baboon Names field identifies individual baboons in the photo by their sname. A drop-down list of all baboon snames was populated from the snames in the Babase BIOGRAPH table (this list will need to be regularly updated). More than one sname may be associated with a single photo.

Behavior: The Behavior field is used to record behaviors captured in the photograph. Possible entries in the drop-down list include: agonism, consorting, drinking, feeding, and grooming.

Location: The Location field identifies where the photo was taken. Possible entries in the drop down list include: Duke, Kenya, and Princeton.

Copyright: The Copyright field is used to record information on who took the photo. Initials are used if the photo credits can be attributed to a single individual (for example, "JA" for Jeanne Altmann and "SCA" for Susan Alberts). "Field staff" should be used if the photo was taken by one of the Kenyan Team members.

Description: The Description field is used for additional information *not otherwise captured in the photo fields*. The following are commonly noted in this field: type of animal (if not a baboon), plant name, Mount Kilimanjaro, vehicle, dead infant, darting, water hole, and tree grove.

By referencing the field staff's file name and personally reviewing the photo subject, fill in as much information in each of the above fields as possible. Ignore any fields that are not applicable to a particular photo.

Renaming Photos
Once all possible information has been entered in the appropriate field(s), you are ready to rename the images. The naming convention for digital photos is a combination of date and sequential numbering in the following format:

$$Y Y Y Y – M M – D D \_ N N N$$

YYYY = Four digits denoting the year. Use 1999 for 1999, 2000 for 2000, 2001 for 2001, etc.

MM = Two digits denoting the month. Always use a leading zero for months with MM less than 10. Use 01 for January, 02 for February, 03 for March, etc.

DD = Two digits denoting the day. Always use a leading zero for months with MM less than 10. Use 01 for January, 02 for February, 03 for March, etc.

NNN = Three digits denoting the sequential numbering of the photos on a particular day. Always use a leading zero for months with numbers less than 100. Use 001 for the first photo, 002 for the second photo, 003 for the third photo, etc.

For example, the first three photos taken on 25 September 2005 are named 2005-09-25_001, 2005-09-25_002, and 2005-09-25_003. This batch renaming can easily be done in Portfolio (refer to the Portfolio User's Guide for specific instructions). If the photo's date was not clearly stated in the filename given by the field staff or the folder it was stored in on the original CD, refer to the imbedded photo data for this information.

Note that ideally the sequential numbering of photos on a given day would be in order of the time pictures were taken. However, this isn't always possible to deduce and occasionally photos for a particular date are added piecemeal to the system. Use your best judgment in numbering the photos and try to avoid renaming photos numerous times.

Archiving Digital Photographs
1. The original photo CDs sent from the field are stored in the Princeton lab. In addition, monthly back-ups of the ABRP Photo file folder are stored on an external hard drive as part of the archiving procedure (see the Back-up and Archive section for specific instructions).

Protocol contributed by Oriana Poindexter

For Color Slides (Beginning in June of 1963)

There is not enough data recorded to name the files with the
Y Y Y Y – M M – D D _ N N N
dating protocol. When naming scanned color slides, use the following:
Y Y Y Y – M M – R R R – N N N

YYYY = Four digits denoting the year. Use 1999 for 1999, 2000 for 2000, 2001 for 2001, etc.

MM = Two digits denoting the month. Always use a leading zero for months with MM less than 10. Use 01 for January, 02 for February, 03 for March, etc.

RRR = Three digits denoting the roll number. Use a leading zero for roll numbers with RRR less than 100. Use 001 for Roll 1, 010 for Roll 10, etc.

NNN = Three digits denoting slide number. Use a leading zero for slide numbers with NNN less than 100, Use 001 for slide 1 of a roll, 010 for slide 10 of a roll, etc. Do not expect to encounter slide numbers above 037 or 038—the three-digit NNN is only being used in order to stay as close as possible to the naming protocol used for the other digital files.

Follow the same procedure as above for the description of the files, entering all information recorded on the slides themselves and all information recorded in the photo log book in the Properties tab.

SEX SKIN SWELLING AND PCS COLOR DATA

The Sexskins table includes information on the swelling size for mature (or nearly mature) females. PCS color data is found in the Pcskins table. In future a view that incorporates the two pieces of data into one record would be desirable. This would also allow for them to be uploaded together, along with any accompanying reproductive notes, and would decrease the amount of manipulation required in Excel to upload the two columns of data separately.

Templates for entry into Excel are created from querying the census table since data on swelling is generally recorded for each day the animal was censused as present. This of course must be done after the demography update and decreases data entry errors. However, adjustments will still need to be made, including adding dates for reproductive data collected on incomplete census days, deleting dates when an animal's reproductive data was not collected on that day. Additionally, prepubescent baboons may need to be removed as sex skin data cannot be uploaded to Babase until after they have had their first cycle. As a result, some baboons may need to be held over until the following update and others will never be uploaded because they died before achieving pubescence. The Sexskins and Pcskins tables are typically updated on an annual basis as time allows for entry and proofing of this bulky dataset. Most sex skin data prior to 1999 are not currently available in Babase but older data will be backfilled as opportunity allows.

Getting Started – Creating an Entry Template
Refer to Appendix 4 for the specific queries used to create the entry template and the logic behind them.

Entry
When doing the data entry, keep in mind that *all* data recorded on the sex-skin sheet should be entered. Also be certain to check for rows either missing or mistakenly added to the entry spreadsheet – these may reflect errors either in Babase or in the field notes. If you discover a row missing in the entry spreadsheet, simply add in that row and enter the data. In the notes field, mark that this row was added during entry (that will flag the proofer to further investigate the problem). Likewise, add a note if you discover a row in the entry spreadsheet for which there is no size or PCS records.

> *Size* – This column contains a number indicating the size of the sex skin swelling. Possible values range from 0 through 20, inclusive. Note that all swelling sizes within this range are intergers with the exception of 0.5, which is used for very small swellings

listed on the datasheets as "0-1". Leave this column blank if no swelling size was recorded that day.

PCS – This column contains letter codes indicating the color of a female's paracallosal skin. Possible values include "P" for pink, "B" for black, and "C" for combined pink and black (indicated as "P/B" on the datasheets). Leave this column blank if no PCS color was recorded that day.

Notes – This column records verbatim any reproductive notes associated with a particular row. Reproductive notes are entered exactly as they appear in the field notebooks (except for cases of spelling/grammar correction or confirmed data corrections).

MAKING CORRECTIONS TO BABASE

Adding PIDs

To correct for a skipped PID in the original entry of pregnancy data, the most important rule to remember is that you should *only change the parity associated with already entered PIDs – do not change the PIDs!*

1. Create a new row in BIOGRAPH for the new infant/abortion (follow the same procedure for this as used in the demography update).
2. Assign the next PID in the series for the mother to this birth. For example, if the mother was SIS and she had 8 PIDs entered to date, this pregnancy would be assigned SIS9. This is true regardless of the pregnancy's date!
3. Reorder the parity numbers associated with each PID for the mother. These PID changes should be made in both the BIOGRAPH and the Pregs tables.
4. Finally, add a row to the Pregs table for the new PID with the conceive and resume cycles entered in.

Deleting a preg that occurred prior to another preg

If an "old" pregnancy is determined via hormone data to have been a shutdown/delay following the resume cycle, rather than a pregnancy and needs to be deleted though there are pregnancies after, you get errors in a loop unless you use the begin and commit commands to make all the changes at once and circumvent the errors. You must also, of course, empty repstats and perhaps the other repro tables in order to not get errors from those. Example: After COB9 was born hormones indicated that Cobra was not pregnant in July 2012 so COB8 needed to be deleted.

```
begin;
update pregs set resume = NULL where pid = 'COB8';
delete from biograph where pid = 'COB8';
delete from pregs where pid = 'COB8';
commit;
```

Updating the GROUPS Table

The groups table defines the group ID (gid) for and parentage of each group and provides information on when each group began, fissioned, was dropped, etc. The table also provides a three_letter_code and a one_letter_code for each group that shows up in GPS or Psion data. The three_letter_code is used to identify groups in focal sampling (Psion) and for other groups sightings recorded in the GPS. The one_letter_code is used to identify focal groups in the GPS data.

When a group fissions (or when two groups fuse), a row is created in the groups table for each new group. The gid should reflect the parentage of the group with the new groups sharing the initial portion of the parent group's gid with an additional decimal added (e.g., Nyayo's group with gid 1.1 fissioned into Hokey's and Snap's groups with gids equal to 1.11 and 1.12, respectively. The gid of the parent group should also be recorded in the from_grp column.

Each group will have one or more dates associated with it. Typically a group will have a start date indicating when the group began. This should always be filled in when creating new rows for fission/fusion products. Most groups also have a permanent date. For the oldest groups, this date indicates when we began monitoring them. For fission products it indicates when the group became an independent, fairly settled group. That is, there may still be floaters but group membership is largely settled and the groups are generally apart.

Parent groups will also have a cease_to_exist date indicating when that group no longer appears in the database because all data is now under the fission or fusion products. Because it is not always clear when a group should be considered to have fissioned, this date may not exactly correspond with the end of a fission, nor even with the last date on which the fission products were together as though they were one group. Sometimes other data connected with the fission products is still listed under the parent group past the last date on which the group was together and data for the group cannot be entered past the cease_to_exist date.

Finally, there is now a last_reg_census date. For groups that have been dropped, this date indicates the last time the group was censused in the regular manner (i.e., using regular group census sheets and collecting the normal complement of other data, as opposed to collecting data in the style of an other groups census). As always, users should keep in mind quirks in the database, like the TZ loss period (roughly 1990-1996) during which group censuses were collected much less frequently than normal. Proton's group's last_regular_census occurs during this period. For groups that have fissioned or fused, last_reg_census indicates the last date on which the groups were censused as the parent group (i.e., the sheets have the name of the parent group) unless there is an obvious earlier point at which to put the date. For practical purposes

this means that for older fissions the date is set as the last date the groups were censused as the parent group but for more recent fissions the date may be set earlier, based on the information available to the database manager.

Nonstudy groups clearly will not get last_reg_census dates and generally will not get cease_to_exist dates, unless the timing of a fission or other group-altering event is relatively well known (as in the case of Ositeti group).

UPDATING THE SNAME LIST FOR AMBOSELI

In an effort to avoid all the complications associated with naming a new study animal a previously assigned sname (*remember that snames are unique!*), the field Team periodically requests an updated list of all the snames that have ever been used as part of the ABRP. This complete sname list comes from combining two data tables: (1) the Babase BIOGRAPH table and (2) a list of project snames that are *not* in BIOGRAPH. The list of snames not in biograph (SNAMES_NOT_IN_BIOG) currently resides in babase_pending. These two sources of snames are mutually exclusive – no sname appearing in one table also appears in the other. Further, the combination of these two lists should be complete – by combining all the snames from both sources, we should generate the complete list of all snames ever assigned to animals in the history of the project.

The querying for snames in the BIOGRAPH table is simple to do. Verify the exact columns that the Team would like to have and simply extract relevant information for all individuals in BIOGRAPH with the exception of fetal losses. Critical data columns usually include: sname, name, birth, matgrp, sex, and statdate. Once this query has been performed, export the result and append it to the table SNAMES_NOT_IN_BIOG.

The origin of the snames not in BIOGRAPH are varied, but fall into four categories:

1. Offspring that were born after we stopped studying a given group. For instance I think there are some Lodge group babies and some Proton's group babies that got named after we stopped monitoring the groups and were not initially incorporated into Babase. These will be added to BIOGRAPH when demography note backfill is completed for these groups.

2. Animals that occurred in the very early days of observation in the late 1960's and early 1970's (or early 1980's for Lodge group), but for whom we have almost no information – for instance BAN who was a putative Lodge group male that occurred there just when we first started but then disappeared. He is not in Babase but we don't want to re-use the name. I believe that BIG (A Hightail's animal) and Boxer and Hairpintail are also in this category.

3. Animals that got named in non-study groups but were never in study groups. DON was an Olkenya female. A number of the others are males in Stud's or Olkenya's that never

immigrated into study groups and that we only knew for a short time. Many of these were added to BIOGRAPH during the demography note backfill effort.

4. Three-letter codes that are used to identify human samples that were used for DNA extraction and PCR, in some cases as human controls and in some cases as checks for contamination. In Susan's lab at Duke, these three-letter codes appear as snames in the genetics database. These names should *not* be used for baboons (they are not very user-friendly so they would probably not get used anyway, but by including them in the list we are more certain to avoid duplicating them with names in the field).

DIRECTORY ORGANIZATION

C:\Database Files\ALTMANN\Babase\
All ongoing production data and programs related to the Amboseli baboon studies are stored in this directory. Special projects, temporary data, and the like should be kept elsewhere.

C:\Database Files\ALTMANN\Babase\DATA
All authoritative computer-encoded data related to the Amboseli baboon studies are stored in this directory. The documentation on each of the databases in this directory, along with descriptions of the data elements, can be found in the *Amboseli Baboon Project: Data Management System.*

C:\Database Files\ALTMANN\Babase\PROGRAMS\
All production programs used in the entry, maintenance, and analysis of the Amboseli baboon studies are in this directory. Documentation on the use of these programs can be found in the *Amboseli Baboon Project: Data Management System.* This also contains the coding standards and design philosophy of the system, which should be followed by anyone modifying or adding programs to this directory.

C:\Database Files\ALTMANN\Babase\Programs 2.0\
Includes programs used for Babase 2.0 entry and analysis. Currently this contains only Ranker.

C:\Database Files\ALTMANN\Babase\TEMP\
All temporary files related to the Amboseli baboon studies go in this directory. This includes all files for upload to Babase. Folders currently in active use at Princeton include CENSUS, REPRO, Ranks, and Meteorological. CENSUS includes the census upload files and demography notes files. REPRO includes files for female cycling, pregnancies, births, maturities, and male dispersals. Ranks contains female rank data files. Meterological currently contains the min_max data (for now WeatherHawk lives in the New Bababase Datasets folder - see below). Upload Logs contains the upload logs from N. Learn. Upload logs for T. Fenn can be found under C:\Database Files\Tabby\update essentials\Demogup_logs.

C:\Database Files\ALTMANN\Data from Amboseli\
This is the warehouse for all electronic data files sent from Amboseli and archived at Princeton. This includes daily/weekly WeatherHawk files, monthly reports, cash accounts, salary calculations, and agonistic matrices from the field Team. These data are organized by

year_month. NOTE: In 2012 the Princeton database manager ceased to receive cash accounts and salary calculation files so these are no longer archived with the other data from Amboseli. Admin assistants at both Duke and Princeton now receive and store these files.

C:\Database Files\ALTMANN\New Babase Datasets\
Datasets currently in development at Princeton for upload to Babase are located here and old files may remain archived here after the datasets go "live" to Babase. These datasets can often be found in babase_pending.

C:\Database Files\ALTMANN\Protocols\
Contains the most current version of this protocol, as well as older versions and assorted other protocols used during the lifetime of the Amboseli Baboon Research Project.

APPENDIX 1: DIRECTIONS FOR COMPLETING THE DEMOGRAPHY UPDATE SUMMARY SHEETS

Pg1: New Individuals (Births, Aborts, Immigrant Males)

Births

Record the <u>infant sname</u>, <u>infant name</u>, <u>birthdate</u>, infant <u>sex</u>, and <u>matgrp</u> for each new birth. The first three letters of the <u>pid</u> can also be entered since they are the mom's sname, but see below for assigning the number portion of the pid. New births should be listed in the field monitoring notebooks in both the Demog and Repro Notes. You should have already cross-checked the sname, name, and sex of each infant among the demog notes, repro notes, monthly report, and neonatal sheets during the monthly data proofing. Refer to Jeanne's notes on the neonatal assessment sheets and sex skin sheets (dates on both sheets should match), to obtain the birth date for each infant.

Important! If a birth occurred at the very end of the last update period, it may not have been entered in the last update. Be sure to check your previous Summary Sheets for these instances, and look at the sex skins as well. Births that fall after the last census day of the update period cannot be entered due to validation rules in Babase. If you find one of these situations in the current update period, it is advisable to copy the birth onto your Summary sheet for the next update period, or at least put a sticky note in the binder to remind you when to enter it.

The <u>bstatus</u> for each birth will be "0" except in the most unusual cases (0 means "known within a few days").

The <u>statdate</u> needs to be temporarily entered as the birthdate for the purposes of the upload. It will auto-update in Babase 2.0 as soon as the new census data are uploaded.

The <u>status</u> for each infant must be "0" (alive) or "1" (dead). See Appendix 2 at the end of this document to assign the appropriate <u>dcause</u> code if your individual has died.

Note: Enter the true <u>status</u> and <u>dcause</u> onto the update sheets, but be aware that in order to upload new births, you have to temporarily assign a <u>status</u> and <u>dcause</u> of "0", "0" respectively, if that infant's sname shows up on the census sheet anywhere (<u>status</u> and <u>dcause</u> cannot be 'null' in BIOGRAPH, otherwise we would just input them as such instead of '0'). If you try to upload census information for a status = 1 (dead) individual, the program complains that a dead individual cannot be censused. This is because the program makes no distinction whether the individual has a census of "1" (absent) or "0" (present). Dead individuals are frequently censused as "Absent" for several days in order to calculate a death date. This is an issue that Karl can probably fix and one which only applies to new births, not immigrant males. (He can probably allow status = '1' to upload IF the census value = A.)   This protocol will later remind you to change the status of all dead infants to '1' later when the program won't gripe at you.

Finally, assign the infant's full <u>pid</u> by querying PREGS and BIOGRAPH (explained below).

PREGS – In most cases, during the previous demography update you should have already assigned a pid to the new birth you are now entering. This was done when the conceptive cycle for that pregnancy was first recorded in Babase. To fill in the pid on the update sheet, look up the mother's pregnancies in PREGS using the following query (ABB is given as an example below):

```
select * from pregs where pid like 'ABB%' order by parity
```

The last entry resulting from the query above should **not** have a resume cycle entered. This is the pid you want. It should correspond to the new infant you are documenting.

BIOGRAPH - Look up all the mother's offspring in BIOGRAPH to make sure the pid you got from the PREGS table is correct. When you run the query below in BIOGRAPH, you should see a pid that is one less than the one you are about to assign.

```
select * from biograph where pid like 'ABB%' order by parity
```

Caution! One possible oddity to be aware of is when the conception and birth occur in the same demography update time interval. Baboons have a gestation period of only 6 months so this typically happens once or twice in every update. In this case, there will not be any pids in the PREGS table that are missing a resume cycle because the cycle for the pregnancy hasn't yet been added to the database! Enter the information to the handwritten summary sheet but refer to the special section at the end of this appendix (When Conception and Birth Occur in the Same Update Period) for instructions on uploading it correctly:

Caution! Be particularly cautious when recording the parity. Parity typically matches the number portion of the pid, however exceptions do exist! Parity should reflect the true chronological birth sequence of offspring. **Parity can be changed in the database if pregnancies are missed and need to be entered years or decades later.** Pid, however, reflects the order in which the pregnancies were input to Babase. **Once a PID is assigned and the demography update complete, it should NEVER be changed.** Many Babase links depend on pid remaining consistent through time.

Aborts

All aborts should be recorded by Jeanne in the sex skin scoring. In this section of the demography summary update sheet:

Enter the pid (follow the above protocols used for births)
Enter the abortdate JA assigned (this eventually goes into BIOGRAPH under birthdate)
Enter the bstatus (generally 0 if known within a few days)
Enter the sex as U (rare exceptions include abortion of a near-full term fetus that the Team happens to see and for which they are able to record the sex)
Enter the matgrp
Enter the statdate (same as the abortdate which = birthdate. The statdate = death date)
Enter the status as 1 (this individual is dead)
Enter the dcause (refer to Appendix 2 for determining dcause, but 7 is common for aborts)

Notes
- Aborts should not have an sname or name.
- The fetus of a female who dies pregnant is treated as an abort.
- Unlike individuals in the "Births" section, you will never have to change the status or dcause in order to upload aborts because they should not have snames and should never appear in a census.
- Aborts should not have an mdate entered for resume in mtd_cycles even though the Team may record bleeding (the abortdate is technically the mdate the same way a birthdate is technically the mdate for the resume cycle after a successful pregnancy.)

Immigrant Males

At the bottom of page 1 is a section for immigrant males (males that were not born into a study group but were first observed/named during the current demography update period).

Enter the sname and name the Team has assigned (from the demography notes).
Enter the pid as NULL (by definition, we don't know what matgrp an immigrant is from)

The Team should prepare an 'age estimation sheet' and a 'hybridity data sheet' for new immigrant males (you should be checking for this during monthly proofing). Use the date of data collection along with the final age estimate and accuracy estimate to assign birthdate and bstatus. For example, if Zibet was estimated to be 10 years old on 25 August 2004 with an accuracy of 1 year, Zibet's estimated birthdate is 25 August 1994 and his bstatus is 1. Be sure to note his assigned birthdate on the demog note and estimate sheet.

For juvenile males, the Team will estimate age in the demography notes by comparing the new individual to an existing juvenile individual. Use the birth of the reference individuals as the birthdate and assign a bstatus of "1" unless Jeanne or Susan indicates otherwise. If the size is estimated as being between two known individuals, split the difference in their ages to determine a birthdate. Also check to make sure the male shows up on the scrotal development sheet for the group the following month. This is what JA wrote about the assignment of bstatus = '1'.

> The reason I think we should assume bstatus = 1 is that size changes so rapidly for males during the juv and subadult period that I think this is safe -- it is hard to imagine that we could assign an age of 5, for instance, and be two years off except in very unusual cases. This is especially true if the Team has been using "1" to mean a window of 2 years, rather than a window of 1 year and "2" to mean a window of four years. [Note: Susan established that the Team was not using these longer windows in March 2011; they are doing it as it is laid out in the monitoring guide, though it was not entirely clear if they have always done it this way.] I still need to discuss this with them, as I am doubtful about this but in the meantime I think we have to go with this.

Enter sex as M.

Enter matgrp as 9.0 ("unknown")
For statdate, enter the birthdate. Statdate will auto-update when you load the census data.

For MOST immigrant males
>   status = 0
>   dcause = 0

However, it is possible that the Team may document the death or corpse of an immigrant male during the same 6 month update period when he joins a group. Under these circumstances, cross out the pre-filled data on the update sheet and replace the status and dcause with appropriate values.

You do not have to fill in the from group column; it is never uploaded into Babase. It's useful only to distinguish *true* new males from those who came from another group.

Pg2: Repro

You might find it easier to complete this page last of all, after filling out pages 4-5: the mtd cycles of each female. The repro section is deliberately printed on pg 2 so it is not on the back of one of the mtd cycles pages. This makes it easier to copy the data from the mtd cycles onto on this summary repro page.

*Remember the following when interpreting Jeanne's sex skin notes:

T	=	tdate		D	=	ddate
R	=	tdate (resume cycle after ppa)		Z	=	ddate
P	=	tdate (puberty)				

Conceptions, Resumes, Puberty Dates (Z, R, P dates)

From the scored sex skin datasheets or pp 4-5, copy the female sname and the P, R, Z dates into the appropriate tables on pg 2 of the update sheets. These P, R, Z data are gathered together on this page because they signify a major change in reproductive state and we need to capture this information.

Conceptions (Zdates)

For every female that Jeanne has marked with a "Z", assign a pid and parity for these new conceptions (Z-dates). Use the following query (ABB is given as an example):

```
select * from pregs where pid like  'ABB%' order by parity
```

This returns a list of all the births to date for that mother. A resume value (i.e. the tcpid value from the MTD_CYCLES view) for the last pid should be listed unless the resume cycle falls in the 6 month update periods that you are currently entering. (Obviously such a resume cycle won't show up until it is entered.)

Assign the new pregnancy as the next <u>pid</u> and <u>parity</u> in the sequence for that mother.   To double-check the <u>pid</u> you assigned, you can also look up the mother's kids in BIOGRAPH using the following query:

```
select * from biograph where pid like 'ABB%'  order by pid
```

A birth for the previous PID should be listed unless during the 6 month update period:
1) an abortion occurred for that PID but the mother quickly became pregnant again
2) the infant was born with that PID but died and the mother became pregnant again

For now you must leave the dcpid column blank on the update datasheet. The dcpid is not generated until the cycles data are uploaded into the database so you must wait to complete this column.

Resumes (R dates)
Review the sex skin sheets or pp 4-5 and enter the <u>sname</u> and <u>Rdate</u> for every female that Jeanne has marked with an "R". You will not ultimately upload any data from this section because Babase 2.0 automatically fills in the resume values in PREGS whenever a new cycle occurs after a conceptive cycle.   The section of the page just helps track which Resume belongs to which pregnancy and that you have the necessary resume cycles post-birth/pre-conception. Babase won't let you load a new pregnancy until the prior pregnancy has a birth in BIOGRAPH and a resume cycle in PREGS.

<u>Pid</u> and <u>parity</u> refer to the pregnancy for which this is a resume cycle (in other words, the last pregnancy). You can query this in PREGS to complete this table.

<u>Conceive (cpid)</u> is listed only to help keep track of cycles. This may be helpful to fill in if this female's resume cycle is also a conceptive cycle (because you won't be able to query it). In this case, the sname should already be entered in the Zdates section above.   Remember Dcpid (of MTD_CYCLES) = conceive (of PREGS) = cpid (of CYCPOINTS). This id is automatically assigned by Babase as part of the upload of Zdates to the PREGS table.

Puberty–female (Pdates)
Copy the "P" dates from the sex skin sheets or from pp 4-5. Enter the <u>sname</u>, and fill in the <u>matured</u> column with the tdate Jeanne has assigned. <u>Mstatus</u> should almost always be "O" (for 'on') unless there was a significant gap in monitoring, in which case it may be appropriate to use "B" (by). Always discuss any unusual circumstances with Jeanne. You can also put a "P" in the <u>P (puberty)</u> column although it is also OK to leave it blank. This table should all be puberty dates so filling it in is redundant.

Puberty-male (Pdates)
Matured "On" Dates for Natal and Immigrant Males
The date of scrotal enlargement is considered the matured "on" date for males, however this is not the same milestone as a male reaching adult status (the latter has to do with winning

agonistic interactions….see the Duke protocol for more details on this). The protocols for determining matured "by" dates, when applicable, are different.

During monthly proofing, you should have already highlighted and flagged the first observation where the scrotum was observed as "E" for enlarged. From the scrotal sheets, enter the <u>sname</u> and the first of the month into the <u>matured</u> column as the matured date for that individual. For example, the first "E" on Lobo's scrotum scoring sheet was marked for June 2004, so 1 June 2004 is recorded under <u>matured</u> for Lobo.

Responsibilities transferred from Duke as of 2013:
You should also assign matured by dates to any new immigrant males who have clearly reached maturity before joining (set it to the first of the month in which they joined) and to any known males residing outside the study groups who do not yet have a maturedate in the month in which they attain the age of 6yrs, 8mos. If the immigrant male is described as a juvenile, he will be added to the scrotal watch list and will be assigned a matured on date when his scrotum enlarges.

You should also assign ranked by dates to any new immigrant males who are clearly adults. These include older males and males that obviously act like adult males in the month in which they join. If you are uncertain (for example if the male is estimated to be about 8yo but is shy or does not appear to have any wins over other adult males, is not consorting, etc.) do not assign a ranked by date; in these cases a rankdate will be assigned at Duke when the male ranks are determined. This is done only in the Excel worksheet and is not noted on the update sheet. Note that ranked on dates are still determined at Duke when male ranks are assigned.

Matured by and ranked by dates are also assigned to natal males who disperse to nonstudy groups or whose group is dropped prior to their attainment of the marker. Unless there is convincing data indicating an earlier achievement of these markers, then the by date is assigned at the median age, as indicated in the rules below. Note that for some years these rules indicated that the median male age of maturity was 6 years, 8months but this was incorrect and ought to have been 5years, 8months. This was discovered on 26 Oct 2015 at which time it was decided that matured by dates incorrectly assigned at 6.7yo would be adjusted to the more recently determined median of 5years, 5months (as recorded in Onyango et al. 2013). Corrections and new assignments for eligible males in nonstudy groups between the ages of 5.41yo and 6.7yo were made in Nov 2015.

Rules for assigning ON and BY dates in various cases are as follows:

1. If the male is natal to our study groups and the date on which the male attained the given maturational milestone is known, he receives an "ON" designation (we know he matured or ranked ON that date).

2. If the male is an immigrant male not natal to our study groups. We follow these rules:

a. If the male enters as a juvenile (field notes indicate testes not enlarged, or notes otherwise indicate that he is juvenile) and stays throughout the entire period of his maturation to adulthood, he gets no "BY" dates. He gets added to the scrotal development sheet automatically (this happens in the field and has been in place for many years) and when his testes enlarge he gets an

"ON" date for MATURED. Similarly, if he goes on to attain rank in the group he gets an "ON" date for RANKED.

b. If the male enters as a subadult (testes enlarged but field notes say he is subadult and he is losing to all adult males in agonistic encounters) he gets a "BY" date for MATURED that is equal to his immigration date. If he goes on to attain rank in the group, he gets an "ON" date for RANKED.

c. If the male enters as an adult (field notes indicate adult) OR he immediately starts winning fights with other adult males, his "BY" date for RANKED and for MATURED is the first of the month he is first censused in a study group.

3. If the male is a natal male from one of our study groups and he disperses before rank attainment, directly into another study group (or after some time alone, but without being in a non-study group for more than a few days). Upon immigration in a study group, he may start winning fights with adult males. This is a common occurrence. We follow this rule:

a. Assign the male a ranked "ON" date that equals the first day of the month in which he immigrated into the non-natal study group.

4. If the male is a natal male from one of our study groups and he disperses around 7 years of age but before rank attainment, but is away from our observations and in an unknown location for a long time (more than a few weeks). We follow this rule:

a. Assign the male a ranked "BY" date that equals the first day of the month in which he first appears as a resident of the non-natal study group. (REVISED 30 Oct 2013)

5. If the male is a natal male from one of our study groups and attained one or both markers in his natal group, but we did not observe him attaining one or both markers (he was already subadult or adult when we started collecting data on him, or we do not have enough data to estimate his dates accurately because, for example, we were unable to observe the group frequently enough during that time -- this happened occasionally, primarily associated with the fission of Alto's), we follow these rules:

a. We assign a matured "BY" date that on the first of the month in which he reaches 5 years 5 months after his birth (this is the median age for testicular enlargement according to Onyango et al. 2013). The point of this is that it will allow us to get a reasonable count of subadult males on any given day in the group, even if the male is not strictly subadult by our definition.

b. We assign a ranked "BY" date on the first of the month in which he reaches 7 years 5 months after his birth (this is the median age for rank attainment according to Alberts and Altmann 1995; note that the median in Onyango et al. 2013 is very similar at 7 years 5.5 months). This likewise allows us to obtain a reasonable count of adult males on any given day in the group.

6. If the male is a natal male from one of our study groups but he disperses before attaining one or both markers, and attains one or both markers in a non-study group. We follow these rules:

a. If he emigrates before testicular enlargement and is known to be living in a non-study group and we have no information about him that would trigger assignment of an "ON" date, we assign him a matured "BY" date that is 5 years 5 months after his birth. Again, the point here is to allow us to (somewhat coarsely) designate the subadults in the population at any given time. This also provides a "BY" date if he comes back into a study group after he has attained the marker.

b. If he emigrates after testicular enlargement but before rank attainment, we assign a ranked "BY" date in one of two ways.

i. If he emigrates before 7 years of age, we assign his ranked "BY" date as 7 years 5 months of age (the median age for rank attainment according to Alberts & Altmann 1995), if he stays in a non-study group that long and we don't have any other information about him. The point here is to be able to designate adults versus other age classes in social groups at any given time. We might also use information in other groups notes, concerning agonistic interactions seen in the non-study group, to give us clues about assigning a "BY" date prior to the median age.

ii. If he emigrates after 7 years of age, we assign his ranked "BY" date as the date he enters the non-study group IF he remains in that group for at least several months. If a male enters a non-study group after the age of 7 years but leaves within a few days or weeks, we do not assume a ranked "BY" date until he enters a group and stays there for some time. This is based on SCA's observation that subadult males on the verge of adulthood tend to stay in a group only when they are successful at attaining rank in that group. Subadults on the verge of adulthood may sometimes "shop around" but they usually leave groups quickly if they are not successful at getting adult rank.

7. If the male is a natal male from one of our study groups but we drop the study group before the male matures or attains rank then we follow these rules: ADDED 30 Oct 2013

a. If the male has already attained the average maturation age (5 years 5 months) when the group is dropped but he is known to remain in the group beyond that time, then he receives a matured "BY" date on the first of the month following the month in which the group is dropped.

b. If the male has not attained the average maturation age (5 years 5 months) when the group is dropped but is known to remain in the group through that age, then he receives a matured "BY" date on the first of the month in which he reaches age 5 years 5 months.

c. If the male is over the age of 7 years 5 months and has not yet attained adult rank when the group is dropped but he is known to remain in the group beyond that time, then he receives a ranked "BY" date on the first of the month following the month in which the group is dropped.

d. If the male has not yet reached the age of 7 years 5 months and has not attained adult rank when the group is dropped but he is known to remain in the group beyond the age of 7 years 5 months, then he receives a ranked "BY" date on the first of the month in which he reaches the age of 7 years 5 months.

Additionally, as of Nov 2015, you will also assign matured by dates to females in dropped study groups when they reach the median maturity age of 4.50 years (as indicated in Onyango et al. 2013) unless data indicate an earlier maturity. Backfilling of such matured by dates was carried out in Nov 2015 (including adjustment of existing matured by dates for dropped females who had previously been assigned matured by dates at the first indication that they had begun cycling). The purpose is similar to that for males – it allows us to get a reasonable count of the number of adult females within the social group on any given date.

As of 2013 you should also assign consortdates for males when they garner their first consort. This is done only in the Excel worksheet and is not noted on the update sheet. Follow these guidelines for assigning consortdates:

With respect to the first consortship date, the criteria that I used were informed very much by the patterns in the data at the time (in the 1990's). Specifically, there were several males who had what I judged to be clearly "false starts" - very short consortships with a long gap after them without any followup (or he disappeared without more activity). The result seems a little arbitrary in retrospect but I think it's reasonable to go with it.

In most of these cases of "false starts" some evidence suggested that they weren't "real" consortships anyway. For instance, several of the cases where a young male was recorded in a very short consortship with a long gap afterwards were cases where it looked like the female may have actually started to deturgesce before the "consortship" began -- that is, she was recorded as turgescent in the morning but her adult male consort partner left her sometime early in the day and she was deturgescent and down in size by the next day. I think this is likely to be a recurring pattern with cases where there is a short consortship record for a young maturing male followed by a long gap with no consorting activity. One of the reasons that we are looking for a consistent pattern of consorting activity after first consortship rather than something that looks like a short one-off.

Males absent from any study group for more than a 2-3 months and receive a ranked by date upon their return do not receive a consortdate, as they could have started consorting while they were out of sight.

Other consortships that technically pass these tests should also be disqualified if they conform to the general idea that it is not a real consortship but just practice with a female who is only recently matured, is not very turgescent yet, or is starting to deturgesce and is not likely to conceive from the consortship in question. In these cases as above there should be no follow up consortships in the months immediately following. Examples include Ralph's only consortship and Luther & Dipstick's first recorded consortships.

Pg3: Deaths
Individuals with prolonged absences (absences at end of month)

Review the census data sheets to find any individuals with prolonged "0" (absent) records. (Look back into the last month of the previous update period to make sure you aren't missing something.) Don't pay attention to the 0's that precede the birth of a new infant or a newly named immigrant male. For all others, record the month, sname, sex, and a brief note on when absent.

Those with a long string of 0's in any month are easy to spot. But the individuals who are seen at the end of one month but simply disappear from the census the next month are more difficult to notice. As part of the protocol for census data update, you will prepare Excel sheets that compile the census information (*Step2b of Date Input/Proofing in Excel* (Enter Census data in a table). While Jeanne is scoring sex skins, you can enter census data into Excel so that you are ready for this section. It's a good idea to make pivot tables from these Excel files and look for

those individuals who may have been present one month, but absent the next (easy to miss if they didn't have a "0" at the end of the first month). Be sure to include these individuals on pg 3 of the handwritten update sheets.

Once you have identified and copied information for individuals with prolonged absences, run the following queries substituting their sname.

```
Select * from biograph where sname = 'EXO';
select * from maturedates where sname = 'EXO';
select * from dispersedates where sname = 'EXO';
```

Enter the birth from the first query onto the summary sheet. Check for a matured date and dispersed date and enter these on the summary sheet, if they exist. You can also check census data especially with the dispersed males. You can do this after you've uploaded the census data for the current update. You may find your male was censused in another group even before he was first censused absent in his own group.

```
SELECT * FROM CENSUS WHERE SNAME = 'LOY' ORDER BY DATE;
```

Deaths

Once you have identified individuals with prolonged absences, determine whether their absence is indicative of a death. Generally, deaths should be assigned only to females and infants/kids (males over four years could have dispersed). Exceptions for adult males include documentations of a corpse or significant, obviously debilitating pathology or injury that the Team has reason to believe led to death. We have also declared very old males who had been resident with a group for over a decade as dead when they have gone missing (e.g., Beam and Alex). Simply disappearing during the same census gap as one or more other baboons with no other evidence to suggest death, is not adequate evidence to declare a male dead age 4 and up and he should be considered censored (which does imply some possibility of death).

If you believe an individual has died, record them in the Deaths section of the sheet entering the sname, birthdate, sex, last census alive, first census dead. Confirm your list of deaths with Jeanne and Susan.

The death date (entered into Babase 2.0 under BIOGRAPH.statdate) should be assigned as the midpoint between the last observation when the individual was alive and the first observation when the individual was missing, unless there is very specific information placing the death on a particular date or indicating a narrower timeframe within which the death might have occurred (e.g., the condition of a corpse could indicate how long an individual has been dead). **The date of the first observation when the individual was missing is included in midpoint calculations.** For example, if a female baboon was observed as present with her group on 8 August and was first observed as absent from the group on 11 August, possible death dates include: 9 August, 10 August, and 11 August. Therefore, the death date assigned as a midpoint value would be 10 August.
Be certain to mark the death date on the field census datasheet in the monitoring notebook with notation similar to:

The <u>status</u> for all dead individuals will be "1" (dead) and <u>dcause</u> should be determined by looking through the monitoring notebooks for data in demography notes, wounds and pathologies, and predation sections. *Refer to* **<u>Appendix 2</u>** *for protocols on determining death dates and notes on assigned death dates. Review all decisions with Jeanne and Susan.*

Dispersal Dates for Natal Males

When a male **with a known matgrp** disappears from his natal group (at any age) and does not already have a dispersedate, a decision must be made about whether to assign a dispersal date to him. This is different than deciding whether he is dead or censored in BIOGRAPH, but it is a related problem and some of the same criteria may apply. Males with matgrp 9.0 (unknown) should not get dispersal dates because their history is unknown.

We use several decision rules, developed by SCA in mid-2011, to decide whether to assign a dispersal date and whether to categorize that dispersal as "confirmed", "very likely", "likely" or "possible" in the dispconfidence column of the DISPERSEDATES table. (The key for the dispconfidence column is found in the CONFIDENCES support table, which is shared by the dcauseconfidence column of the BIOGRAPH table.) We also use a set of decision rules about how to decide when he dispersed. These are addressed separately below.

A. Should he get a dispersal date and how should it be categorized (i.e., what dispconfidence should be assigned to him)?

1. If he was seen either alone or in another group after he left the natal group (see section B for how to determine what "left the natal group" means) then he is a confirmed dispersal and you should assign him a dispersal date and categorize this as confirmed (see section B for how to figure out the date).

2. If he was never seen again after he left the natal group, are there demography notes to indicate that he recently was visiting other groups, spending a day or two alone, or spending a lot of time watching other groups up in trees or on hills?

    2a. If yes and he was in the prime dispersal age group (7 – 10 years old) or the early dispersal age group (4-6 years old) then you should assign him a dispersal date and categorize him as a "very likely" dispersal.

    2b. If yes and he was in the juvenile age group (< 4 years old) then you should assign him a dispersal date and categorize as a "likely" dispersal.

    2c. If no and he was in the prime dispersal age group (7 – 10 years old) then you should assign him a dispersal date and categorize him as a "likely" dispersal.

    2d. If no but he was censused in the natal group during the observation period and then was not there later in the day and there were no signs of any predation event or

illness, then you should assign him a dispersal date and categorize him as "confirmed" if he was in the prime dispersal age group or early dispersal age group, and "likely" if he was in the juvenile age group.

2e. If no and he was in the early dispersal age group (4-6) then you should assign him a dispersal date and categorize him as a "possible" dispersal.

2f. If no and he was in the juvenile age group (<4 years) then you should not assign him a dispersal date (regardless of whether you consider him censored or dead in BIOGRAPH).

B. What dispersal date should be assigned to him?

A natal male is considered to have dispersed from a study group when he is absent from the census for that group of 3 or more consecutive days of observation. If he visits other groups for shorter periods of time but comes back to his natal group, having been missed on only one or two censuses, you should not consider him to have dispersed. If he visits another group and is missing from his natal group for 3 or more consecutive census days, you should consider him to have dispersed even if he comes back to the natal group at some point. It's not perfect but nothing really is.

The specific date of emigration is taken as either the known date on which the male left the group or the midpoint between (i) the last date the male is seen in the group and (ii) the first date the group is seen without the male or the first date the male is seen away from the group, whichever is the shorter interval.

> *Exception*: In some cases, a male transfers from a non-study group to a study group and the non-study group is not seen for some time before the transfer. If the male was known to be in the non-study group the last time the group was seen and if it was seen within approximately one month before his transfer, and if there were no records of the male being seen anywhere else after the last time that the non-study group was seen, then the date of emigration from the non-study group is taken as the day before immigration into study group.

With these rules and exceptions in mind, the steps involved in actually calculating dispersal dates are as follows:

1. Scan through your compiled list of prolonged absences. Don't forget the periods of absence that start and the end of one month and continue through the beginning of the next month.
2. For each case where a male is marked absent for at least three consecutive monitoring days, enter onto the summary sheets the date he was last seen alive and first register absent.
3. Check your Excel census entries for presence in other groups during the time of his absence bout in his natal group.
4. Check demography notes for any evidence that he was wandering, visiting, spending time

alone, or looking into the distance at other groups for extended periods of time.
5. Incorporate any findings from steps 3 and 4 with the absence dates in the natal group to calculate a disperse date (per instructions listed above).

Once you calculate a dispersedate, enter it into the field in the Male Dispersals section of Page 3 and note it on the census page in the binder as well.

Pg 4-5: (M, T, D Cycles)

All information on the cycle dates (mdate, tdate, and ddate) for adult females should be collected onto pages 4-5 of the summary data sheets. Go through the sex skin pages month by month and record the mtd dates for the individual females. If a female doesn't have all 3 dates on a single sex skin page, look ahead to the next month or back to the prior month to make sure you don't have lone m, t, or ddates. The computer will complain about lone dates or automatically generate mdates that create errors if you have a real mdate, but haven't attached it properly to its tdate and ddate in a cycle.

*Remember the following when interpreting Jeanne's sex-skin notes:

$$Z \quad = \quad ddate$$
$$R \quad = \quad tdate$$
$$P \quad = \quad tdate$$

1. Remember that mdate might not always be observed in the field and that mdates marked during a pregnancy or after a tdate should *not* be entered. In the first case, simply leave the mdate field blank. In either of the other instances, do not record any mdate that is "out of order." Also note that it is acceptable for an mdate to equal a tdate. Circle the mdates in red pencil on the sex skin sheets.

2. If a cycle does not have a t and ddate in the current demography update, do not enter it – it is an incomplete cycle! Create a file for the next update and just enter the incomplete (or "dangling") cycles so you don't miss it on the next update.

3. Babase 2.0 will not allow you to enter any data attached to a date that exceeds the statdate for the individual. The statdate is automatically updated with the individual's last census day (or you can manually change it for a death). For example, even if Jeanne has marked a ddate of 31 Dec on a sex skin, if the last census occurred on 29 Dec, that entire cycle cannot be entered until the next update. (This rule also holds for births falling after the last census day). Write the date of the last group census on the top of the Cycles page in the space provided. This will remind you that cycles falling after this date will have to be entered in the next update.

4. Keep in mind that every Zdate you wrote down as part of the pregs data collection is also a ddate for a cycle. Do a quick spot check between the Zdates table and the mtd cycles

you have gathered. Every Z date should have a matching Ddate somewhere. Do the same check between the Rdates table and the Tdate of the mtd cycles.

When Birth, death, and conception Occur in the Same Update Period

When there is a conception, birth or abortion, and subsequent conception in the same update period, the upload process gets tricky. You have to be very careful to do a sequential input for females with multiple conceptions during a 6-month period. You cannot input a new conception in PREGS unless there is a cycle for the conception date. You cannot input a cycle unless there has been a birth from a prior pregnancy. You cannot input a birth unless there is a record of the conception.

Normally we enter the abortion/birthdate first, but the system will throw an error if the individual has a pid that cannot be found in PREGS because s/he was conceived in the last 6 months and that cycle has not yet been uploaded. If the female conceives and aborts (or gives birth) during the same 6-month period, enter the record into BIOGRAPH excluding the pid (you can put a comment next to it in Excel to provide the proper pid to fill in later on). You must do this because BIOGRAPH is the first table you update. You are allowed to do this because it is the same way in which you would enter an immigrant male: it is not required to have pregnancy or conception information on every individual in BIOGRAPH.

Next upload the new cycles and pregnancy data. When you return to BIOGRAPH as part of Step 6 (update the BIOGRAPH table with death information) you can fill in the pid.   Then with the cycles upload you can include the conception cycle for the individual.

If it happens that the infant dies/aborts, and there is another conception quickly, you will have to pull out the record of the second pregnancy (from the PREGS update) and the second conception cycle (from the MTD_CYCLES update) and save them to a separate file. You must go to BIOGRAPH and assign the pid, status and dcause of the aborted/dead infant. Once the first fetus is 'born' in BIOGRAPH, you will be able to input a new cycle and conception. A female can't have another cycle while she is pregnant (i.e. before a resume cycle is listed in PREGS.). She can't resume cycling until that pid is 'born' into BIOGRAPH.

APPENDIX 2: DCAUSES

J. Altmann
Revised: 9 August 2006, 7 January 2011, 2 June 2014

Criteria for assignment of mortality cause, including the range of evidence (strongest followed by most circumstantial) used for assignment to that category.   If a mother and infant disappear at the same time and a cause is attributed to the mother's death, the same cause is attributed to the infant's death unless contraindicated by available evidence.

1. *Predation*: Evidence ranges from (strongest) observed predation or fresh, partially eaten carcass of very recently healthy individual, through (most circumstantial) disappearance, **within four days of observation**, of fully healthy-appearing individual. Intermediate evidence includes group giving alarm barks and other behavior that is tightly associated with predators. In the case of infant mortality, intermediate evidence includes instances in which the infant disappears and the mother is wounded. Wounds presumed to have been inflicted by a predator may not kill the baboon until after the baboon is seen again but deaths resulting from these wounds should still be considered predation.

2. *Conspecific*: This category includes both death resulting from infliction of lethal wounds and death resulting from kidnapping of a young infant that then dies of starvation and dehydration when deprived of its mother's milk. Evidence for kidnapping is direct. Death only occurs if the kidnapping persists for several days and it is, therefore, unlikely to go unrecorded or to occur in any other than infants less than a week or two of age. Infants older than this are mobile enough that kidnappers are unlikely to be able to keep them for more than a few hours. Evidence for lethal wounding ranges from direct observation of the wounding through either appearance of puncture wounds on head or torso of a young infant and absence of wounding on the mother or prior records of intense attacks on infants followed within a very few days by an infant's disappearance (see Palombit 2000, 2002). Because of the intense interest and controversy surrounding the topic of aggressive infanticide and the potential factors that may have selected for the behavior, we have chosen to include and report separately for this category both the fully observed and the circumstantial cases.

3. *Wounds and injuries from accidents*: This category is restricted to wounds or injuries thought to have been derived from accidents. Examples include a broken leg, back injury, or internal injuries that result from a fall from a tree. Evidence can range from witnessing the fall or its aftermath to inferring that a fall occurred based on the type of injury. Wounds caused by encounters with nonpredator mammals, such as gorings by elephants or wildebeests, should also be considered accidents. Previously wounds that could not be attributed to predation or conspecific wounding were sometimes placed here but these should be assigned to dcause 7.

4. *Pathology or obvious congenital problems*: Evidence ranges from (strongest) observed signs of severe illness or weakness through (more circumstantial) high prevalence of illness in the group combined with a gap in observations of 5-14 days during which the individual disappears. For very young infants, signs of illness or weakness that lead to death are often recorded in the neonatal assessment records taken twice in the first week of life. For

prevalent illnesses, our records include several instances of apparent epidemics, some involving paralysis and likely due to coxsackie virus. If an infant dies at an age at which mothers usually carry the corpse for several days or more, lack of major wounds on the corpse is also taken as evidence for placement in this category. Also included in this category are individuals with a severe disability that persists after an injury has otherwise healed.

5. *Loss of mother*: This category includes instances where an infant disappeared simultaneously with or very shortly after its mother and the death is not assignable to any of the other categories. Loss of mother is used for all cases where an infant died in utero upon the death of the mother. It also includes prolonged separation from a live mother (e.g., Moon).

6. *Human action*: This source of mortality has been observed primarily for the groups living around the tourist lodges and during the 2009 drought when human-baboon conflicts over goats occurred. Evidence ranges from finding an individual with machete wounds or injuries obtained from having been hit by a vehicle, usually combined with confirmatory reports or rumors in the lodge worker or the Maasai community, through sudden disappearance combined with reports of an animal being so injured.

7. *Unknown*: This category includes any instances that do not meet any of the criteria above. This category is routinely assigned to fetal losses or stillbirths when the cause of death is unknown (known causes are loss of mother or same dcause as mother). It is also assigned when the animal was known to have a serious wound but the source of the wound is unknown. The unknown category is also used when an animal disappears during a larger gap in observations (greater than four days since last observation).

8. *Under review*: This was set up as a temporary "hold" category for dcauses that have yet to be properly assigned. This category should go away once demography note backfill and concurrent dcause review is complete.

99. C*ensored*: Baboon is no longer under observation and may or may not be alive.

Additional notes:
- The carcass of a dead infant is sometimes carried around by adult(s) for some time after the infant's death. The field Team's notes on the condition of the infant's carcass upon first observation can provide useful cues as to when (and also how) the infant died. These records need to be brought to J. Altmann as the season and other aspects of the carcass may affect the determination of death dates.

- The Team may record when one or more baboons in the group are giving lost calls following the disappearance of another group member. This is a useful clue for determining when an individual disappeared from the group, as lost calls are not typically given for more than 2 days.

- Keep in mind that when a full review of all data available regarding a baboon's death indicates that the death date should be assigned as the midpoint between the last observation when the individual was alive and the first observation when the individual was missing, *the date of the first observation when the individual was missing is included in midpoint calculations*. For example, if a female baboon was observed as present with

her group on 8 August and was first observed as absent from the group on 11 August, possible death dates include: 9 August, 10 August, and 11 August. Therefore, the death date assigned as a midpoint value would be 10 August. *(Note that, in this case, death on 8 August would only be assigned if specific, detailed data on the baboon's death on that exact day existed.)*

In 2011 we also added confidence levels to dcauses and dispersals. These can be found in the CONFIDENCES table in Babase. For dcauses and dispersals the confidences are as follows:

- 0 – Not applicable (used only in the dcause column)
- 1 – Possible, with limited evidence
- 2 – Likely, with circumstantial evidence
- 3 – Very likely, with corroborating evidence
- 4 – Confirmed
- 8 – Under review; this is a temporary category assigned to deaths and unconfirmed dispersals prior to 2011 until those confidences can be backfilled (which will occur concurrently with demography note backfill, now in progress).

Note, however, that 0 is not used for dispersals. See the section ***Dispersal Dates for Natal Males*** above for more information on assigning dispconfidence values.

The confidence level for a dcause is found in the dcauseconfidence column of BIOGRAPH and the confidence level for a dispersal is found in the dispconfidence column of DISPERSEDATES. See the section assigning dispersal dates for information on dispconfidence assignment.

A dcauseconfidence of 0 is assigned when a baboon is still alive, when a baboon is censored, or when the dcause is 7 (unknown) or 8 (under review). If the status is 0 or 2 or if the status is 1 and the dcause 7 or 8, then the dcauseconfidence should always be 0.

A dcauseconfidence of 1 is most often assigned in cases of presumed predation. It may also be assigned in conjunction with other dcauses when we have some slight reason to presume another cause of death but have no substantial evidence to support a stronger claim. Cases with a dcauseconfidence of 1 should be differentiated from other cases when examining dcauses.

A dcauseconfidence of 2 is assigned when there is a small amount of evidence that suggests a specific cause of death. It may also be assigned when evidence potentially supports more than one dcause but we think the evidence leans toward one or the other. For example, if we know a baboon died from a wound and we think the wound was obtained from a conspecific, rather than from a predator (or vice versa), but cannot rule out the alternative, then a dcauseconfidence of 2 may be assigned.

A dcauseconfidence of 3 is assigned when there are multiple lines of corroborating evidence and little doubt that the cause of death could have been otherwise than assigned.

A dcauseconfidence of 4 is assigned when there is no doubt that the baboon has died of this cause. Often in these cases a body is recovered or the death is witnessed or there is overwhelming evidence leading up to the baboon's disappearance that death was imminent.

APPENDIX 3: QUERIES FOR FECAL PREP CLEAN-UP

D. Onderdonk
Last Revised: 27 May 2004

Note: Do queries 1-4 (cleaning up bad snames and names) before doing 5 and 6, since 5 and 6 link Fecal PREP to BIOGRAPH by sname, and so bad snames will drop out of these queries. Have Fecal PREP and BIOGRAPH open in the data session window.

1. Check that all snames in Fecal PREP are also in BIOGRAPH. In addition to bad snames, abbreviations for unnamed individuals will show up. Also any individuals who are not yet in BIOGRAPH will show up.

   select * from prep where sname not in (select sname from biograph) order by sname

2. Check that all names in Fecal PREP are also in BIOGRAPH. As above, abbreviations for unnamed individuals and any individuals not yet in BIOGRAPH will show up here, in addition to the incorrect names. *Note that if the misspelling of the name involves the first 3 letters of the name, be sure to also correct the sname.*

   select * from prep where name not in (select name from biograph) order by name

3. Check that all snames in Fecal PREP are in fact the first 3 letters of the full name. In addition to mistakes, unnamed individuals and OFR (Offer's sname) and DOW (Dhow's sname) will show up.

   select * from prep where sname # left (name, 3) order by sname

4. Check that all of Offer's samples have her correct sname of OFR.

   select * from prep where name = "OFFER" and sname # "OFR"

5. Check for samples where the sample date is before the birth date of the animal. Any individuals with no sample date in Fecal PREP will show up. There will also be lots of blank lines for individuals with no sname or name – ignore these.

   select sid, prep.sname, prep.name, date, birth, status, statdate from prep, biograph where date < birth and prep.sname = biograph.sname

6. Check for samples where the sample date is after the statdate (last date seen) of the individual. You can ignore the ones where the statdate is just the end of the last census update (the query orders the results by statdate to make this easy). These just mean the sample was taken after the end of the last census update (make sure you know when Babase is updated through).

   select sid, prep.sname, prep.name, date, status, statdate from prep, biograph where date > statdate and prep.sname = biograph.sname order by statdate

7. Find any samples where dates do not occur in the relative order that they are supposed to, in order to catch date errors.

    select * from prep where fzdried < date and fzdried # {}

    select * from prep where siftdt < fzdried and siftdt # {}

    select * from prep where meextract < siftdt and meextract # {}

8. Find all non-blank dates for these variables in order – check the beginning and end for funky dates (funky dates in other columns should have been caught with the last set of queries, but they can be checked this way too).

    select * from prep where date # {} order by date

    select * from prep where meextract # {} order by meextract

    select * from prep where spe # {} order by spe

Other useful queries:

To check the proper spelling of a name:

    select * from biograph where sname = "XXX"

To change several similarly misspelled names all at once:

    update prep set name = "NEWNAME" where name = "OLDNAME"

APPENDIX 4: CREATING A SEX SKIN DATA ENTRY TEMPLATE

C. Markham
Revised by Tfenn: 24 April 2009

Objective: The notes which follow provide specific query steps involved in preparing a sex skin data entry template table for females (in this example, the dates included in the template are specifically for females censused from July 2005 through December 2005 for the 05B demography update). The table will include all females after their mature date as well as juveniles 3 years, 8 months and older (the field Team generally begins recording sex skin data when a female is 4 years old – Jeanne would like the datasheets to include dates 4 months prior to the individual's 4th birthday to catch any early observations that may have been made). This approach may capture a few extra snames of adolescents that the Team has not yet started monitoring, but these can easily be deleted from the prepared sex skin table if there is no available information.

The one-step query for preparing a sheet of sex skin snames from Babase 2.0

```
select extract(day from date) as day, extract(month from date) as
month, census.date, census.sname, census.grp from census left
outer join biograph on biograph.sname = census.sname where
census.grp = 1.11 and census.date >= '2014-01-01' and census.date
<= '2014-12-31' and biograph.sex = 'F' and (date-birth)/365.25 >=
4 and census.status <> 'A' order by extract(month from date),
census.sname, extract(day from date);
```

The <> 'A' portion is very helpful when there are floaters following a group split; otherwise lines are generated for the absences as well, which is quite a mess.

Check the output and make sure you are satisfied you have the correct data and group. You can always check your query by looking at the maturedates table to see which females reached puberty during a given year. Just make sure you captured all births for females that would appear in MATUREDATES. You can use a query like this one:

```
select maturedates.*, biograph.birth, biograph.matgrp from
maturedates left outer join biograph on biograph.sname =
maturedates.sname where maturedates.matured > '2008-01-01' and
biograph.sex = 'F';
```

If you think you've got everyone, you can download the results of the first query from PPA and bring it into Excel. Add new columns for **size**, **color**, **color changed** and **notes** (no caps). This is the easiest way for people to do the data entry since all the data are captured on one sex skin sheet. However, for the purpose of upload, you will have to break this one worksheet apart into two separate worksheets that are uploaded to the tables SEXSKINS and PCSKINS. They would look roughly like this:

SEXSKINS			PCS KINS		
sname	date	size	sname	date	color
CHE	5-Jan-99	0	CHE	5-Jan-99	B
CHE	7-Jan-99	0	CHE	7-Jan-99	B
CHE	9-Jan-99	0	CHE	9-Jan-99	B

Remember that for each group, you must change the group number in the query to get the snames for each group. When you want to do a new year, you will have to change the census date range to match your year AND change the birth date to search for females who were more than 3 years, 8 months old at the beginning of you period of interest.

Status of Sex Skin Uploads.

Please see the status table on the wiki for the most current information.

Important Note ***

Juvenile females who are under observation and die while under observation must be deleted from the upload file. They will generate an error message because they have no cycles to which their sex skin size can be related and all records in the database MUST be related to a cycle in order to appear in the sex skin table. (It is not clear if Jeanne wants these individuals deleted from the Excel sheets as well and a note left with their sex skins explaining why they were never entered in the database. Check with Jeanne on this.) Also check with Jeanne regarding juvenile females who have mini-cycles while still in a pre-pubescent phase.

APPENDIX 5: ADDING AND DELETING BABASE ACCOUNTS

Periodically you may need to add new users (create logins) to Babase or purge users who are no longer with the project. This is done using SSH after logging into your Unix account. Please see the instructions on the wiki at http://papio.biology.duke.edu/babasewiki/DataManagement#newuser.

Also, to change your password on a regular account, from the database homepage in papio go to the Account tab along the top and then click "Change password". To change your password on an admin account, instead of the Account tab you will find a Roles tab. Go there and select Alter on the row for your admin account – this will open a page that allows you to change your password.