AMBOSELI BABOON PROJECT

Protocols for Data Management

Altmann Lab – Princeton Work

Updated September 2020

Written and compiled by a series of Princeton data managers
Most recently updated and edited by Niki H. Learn

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 keeping an update log and, when an update is complete, send an announcement on the Babase mailing list and update 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 of incoming data from Amboseli
2. Data checking and entry with questions sent to the field team (the Team)
3. Data proofing and correcting
4. Data uploading and validating (don't forget to check the Warning System)
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 by the Team 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 project 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 on 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 the Alberts lab.

Filing of Photocopies

Each package of data should include a Data Sent List. Inventory the data by sorting through the contents and checking off items on the Data Sent List, verifying the number of pages as you go. 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.* The min_max data are available in Babase and the WeatherHawk data were moved to Babase from babase_pending in Mar 2017. Both are updated quarterly.

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." When the monthly data started being scanned in camp we stopped printing these so more recent copies are only stored electronically with the rest of the monthly data.

Neonatal Assessments – There are two "Neonatal Assessments" notebooks currently used for the offspring of living females: one for groups descended from Alto's group (Dibble's, Vogue'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 at the end of July 2012 and Snap's and Mica's groups ceased to be monitored after Dec 2012, except for occasional other groups censuses. 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 at the end of each 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 as more than one month's data is usually recorded on each sheet – be certain to keep only up-to-date versions 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. HYBRIDITY_RAWMORPHO is updated with these scores during the quarterly Babase update. Related tables in babase_pending – HYBRIDMORPHO (calculated from the raw morphology scores determined in the field) and HYBRIDGENE (derived from genetic data) – are not updated quarterly but might be updated in batches again in future.

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 (which have not been done since 2014). The tables are updated in Excel monthly as new age estimates arrive. Currently these data are not available anywhere in Babase but are kept in an Excel sheet. For a time they were periodically provided to Jutta Gampe for statistical analysis.

Other Group Censuses – 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 Dibble's, Snap's, Mica's, Kelly's (A & B), Ositeti (Jill's and Ceejay's), Olkenya, the Sinya groups (yes, groups – now there are individual designators for several Sinya groups, currently

Sinya A, B1, B2, and C, but for a long time there were two to three groups that were represented by grp 13), Joy's, Nzige's, Proton's (which eventually became one of the Sinya groups), and Stud's. These data are organized by group and date in two "Other Groups" notebooks and the "Dropped Groups" notebook for those groups dropped in 2012 and 2013.

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 residing outside of our current study groups were added to help keep better track of these baboons. These checks are performed by our drivers during the other group census day and/or during the last days when personnel are available. The sheets include space for recording notes on where the baboons were seen and demography notes are derived from these sheets when known individuals are recorded. They are filed in a separate section of the collar checks notebook.

Tree Grove Monitoring – These are historical, noncurrent data and are filed in chronological order in the "Tree Grove Monitoring" notebook.

Daily Activity Calendars – Alphabetize monthly calendars by employee's first name and staple together in one monthly packet. We also receive 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 all now filed electronically and not printed. Additionally in 2015 the observers and drivers began filing activity in a Google spreadsheet so calendars are now received only for the camp support staff.

Amboseli Log Sheets – Beginning in 2006, the 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). Log sheets have been added or dropped as needed. They 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. In the past occasional digital photo CDs were also sent. Refer to the "digital photos" section of this protocol guide for notes on managing and cataloging images. The disks are 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 and in late 2015 we implemented use of Dropbox for transfer of files, including photos, from the field to the US.

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. These are now filed electronically with the monthly data and not printed.

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 eventually arrive at Duke when one of the Principal Investigators travels to Kenya and carries them back. 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 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 get sent to Duke via regular U.S. mail. Originals are still carried to and stored at Duke.

Emailed Files

The Team also emails electronic data files. These files contain three main types of data: meteorological data collected by the WeatherHawk station (formerly.csv, using the old WeatherHawk software, but now .dat using the official VisualWeatherHawk software), point sample data (.pts) which mid-2015 were replaced by Samsung Prim8 data (.csv), and SWERB GPS readings (.mps originally but .txt for many years). The Team also began putting copies of these files into Dropbox folders at around this same time. Then, in late May 2018, the Team also began scanning and sending copies of their reorganized field pages (see next paragraph). All these data are sent to both Princeton and Duke: Princeton works with the meteorological data (sent weekly), while Duke works with the point sample data and the GPS readings (formerly sent weekly but as of 2015 sent daily). Duke also enters the ad lib and mounts and consorts data from the small notebook pages while Princeton is responsible for the usual data integrity checks. The meteorological data are saved 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.

Beginning in late March 2018 we changed the format for collection of mounts and consorts and ad lib data. Formerly agonisms were collected on one centralized set of sheets, groomings on another, mounts and consorts on another and these were added to

throughout the month. Now each observer daily creates their own sheet for ad libs (with groomings and agonisms mixed together), one for mounts and consorts, and one for the focal sampling schedule and any notes (such as demography or other groups notes or predation and human disturbance notes if they cannot immediately put them in the proper place). These sheets are scanned and sent daily along with the GPS and Samsung files. At Princeton they are typically printed daily and filed by group until the monthly data arrive. As they arrive we check for missing pages, incorrect or incomplete dates, or other obvious problems that can be corrected swiftly. Once the monthly data arrive the regular checks against the census dates, female cycling data (for mounts and consorts), swerb data (for mounts and consorts times), and comparison of any notes on the focals pages with what notes entered in their proper location (both to check for missing demographic notes, wounds and pathologies, or predation and human disturbance entries and to look for any conflicting or additional information). These checks are much the same as before only now the data are spread across many more pages. It is more work for the data managers but seems to result in fewer transcribing errors and less missing data than the prior method and the field team likes it.

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. Previously the salary calculation and cash account files were archived in Princeton along with the weekly files mentioned above and sent to Jeanne's assistant for accounting but these things are now handled at Duke. 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, field-entered interaction data, and fecal supply lists may be sent separately as they become available.

FILE NAMING CONVENTIONS FOR THIS DOCUMENT

Whenever a table from Babase is named in this document, it might or might not appear in all caps, such as BIOGRAPH.

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

Specific file names may 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 database, 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**, while weather data resides in **Meteorological**.

There are also many electronic matrices exported from the program *Ranker;* these and other files used in determining ranks 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 and later versions, 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. These logs are filed in C:\Database Files\ALTMANN\Babase\TEMP\Upload Logs.

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**i**nda'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, La**z**a'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,

Pregnancies, 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. **R**anked by dates (for certain males), **cycgaps**, and **consortdates** also have been uploaded at Princeton since partway through 2013.

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 at Princeton 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 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. (Note, however, that later corrections may not show up here.)

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 Groups" 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. **Note:** While the files living in this folder have been used to replace the members-style data for the years included in those files (entered at Duke), there are still many group-years of members-style census data still residing in Babase. Census backfill is not complete and will be proceed hand in hand with demography note 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 (beginning with the 09a update) and death and dispersal confidence levels (introduced sometime in 2011, with dcauses and dcause confidences revised at the start of 2017 to separate the confidence into two parts – dcausenatureconfidence and dcauseagentconfidence), as well as files concerning initial censoring of individuals no longer within the study population (Nov 2009) and backfill of 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 behavior that's not demographic-relevant 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. This index is copied to the Duke database manager each quarter so that misplaced information pertaining to Duke's datasets can be added to the database.

DEMOGRAPHY UPDATES

A "demography update" refers to consolidating the following hard data into computer files that are subsequently uploaded into the Babase database tables and views. It is useful to have at least one month's data beyond the target time period to complete the demography update but since switching to quarterly updates we have generally done without, which means the database manager must be extra vigilant since the final month's pregnancies will generally not be available and some possible pregnancies from the middle month of the quarter still may be uncertain. Note that cycling data will have a wider margin for error at the end of the update period, as repstats and cycstats will be inaccurate 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	MATERNITIES BIRTH_GRP ENTRYDATE_GRP
Aborts	upload program	BIOGRAPH	Table	BIOGRAPH	MATERNITIES BIRTH_GRP
Immigrant M	upload program	BIOGRAPH	Table	BIOGRAPH	BIRTH_GRP
Census	upcen program	CENSUS	Table	CENSUS BIOGRAPH MEMBERS	CENSUS_DEMOG (sorted) DEMOG_CENSUS (sorted)
Groups	SQL Update	GROUPS	Table	GROUPS	GROUPS_HISTORY
Behavior gaps	SQL Update	BEHAVE_GAPS	Table	BEHAVE_GAPS	
F repro cycles	upload program	MTD_CYCLES	View	CYCPOINTS CYCLES PREGS MMINTERVALS* MDINTERVALS* CYCSTATS* REPSTATS*	MTD_CYCLES MATERNITIES CYCLES_SEXSKINS (sorted) SEXSKINS_CYCLES (sorted) CYCPOINTS_CYCLES (sorted) CYCSTATS_GRP* REPSTATS_GRP*
Pregnancies	upload program	PREGS	Table	PREGS	MATERNITIES
Cycgaps	upload program	CYCGAPS	Table	CYCGAPS CYCGAPDAYS CYCSTATS* REPSTATS*	CYCGAPS_GRP CYCGAPDAYS_GRP CYCSTATS_GRP* REPSTATS_GRP*
Maturedates F	upload program	MATUREDATES	Table	MATUREDATES	MATUREDATES_GRP
Maturedates M	upload program	MATUREDATES	Table	MATUREDATES	MATUREDATES_GRP

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 & upload program	BIOGRAPH & Currently in babase_pending	Table & Pending Table	BIOGRAPH & DEATHS	STATDATE_GRP (& none)
Dispersal M	upload program	DISPERSEDATES	Table	DISPERSEDATES	DISPERSEDATES_GRP
Rankdates	upload program	RANKDATES	Table	RANKDATES	RANKDATES_GRP
Consortdates	upload program	CONSORTDATES	Table	CONSORTDATES	CONSORTDATES_GRP
Sex skin size	upload program	SEXSKINS_CYCLES	View	SEXSKINS	CYLCES_SEXSKINS (sorted) SEXSKINS_CYCLES (sorted)
PCS color	upload program	PCSKINS	Table	PCSKINS	PCSKINS_SORTED PCSKINS_GRP
Manual weather data	upload program	MIN_MAXS	View	RAINGUAGES TEMPMAXS TEMPMINS WREADINGS	MIN_MAXS (sorted)
WeatherHawk weather data	upload program	WEATHERHAWK	Table	WEATHERHAWK	
Wounds and pathologies	upload program	WP_REPORTS_OBSERVERS WP_DETAILS_AFFECTED_PARTS WP_HEALS	Table	WP_REPORTS WP_AFFECTEDPARTS WP_HEALUPDATES WP_OBSERVERS	WP_REPORTS_OBSERVERS WP_DETAILS_AFFECTEDPARTS WP_HEALS WOUNDSPATHOLOGIES
Neonatals	upload program	Currently in babase_pending	Pending Table	NEONATALS	
Subgroups	upload program	Currently in babase_pending	Pending Table	SUBGROUPS	
Intergroup encounters	upload program	Currently in babase_pending	Pending Table	INTERGROUP_ENCOUNTERS	
Hybridity	upload program	Currently in babase_pending	Pending Table	HYBRIDITY_RAWMORPHO	
Corpse_info	SQL Update	Currently in babase_pending	Pending Table	CORPSE_INFO	

* Table/view change depends upon running the "rebuild" program for the relevant table.

There should rarely be direct manual input into the tables of Babase tables during these updates other than applying statdate, status, dcause, dcausenatureconfidence, dcauseagentconfidence changes to BIOGRAPH and adding manual census points on exit dates. GROUPS, BEHAVE_GAPS, and CORPSE_INFO are rarely updated and typically are updated outside of the normal quarterly update process. **The update files should always be uploaded to Babase 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 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 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 the Team about uncertain pregnancies, new individual names, and infant sexes as needed

1d: Evaluate disappearances to assign death and dispersal dates and confidences

Step 2 - Date Input/Proofing in Excel

Create and populate Excel tables for (in no particular order)

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

2b: Census data

2c: Matured dates (m and f dates) and ranked by dates where applicable

2d: Male dispersal data

2e: Female cycling data

2f: Pregnancy data

2g: Demography notes (notes on individuals residing in study groups, derived from demog notes or other groups notes in group binders and occasionally from odd sources such as GPS lone animal sightings)

2h: Other groups notes (notes on known individuals living in nonstudy groups, derived from 'Other 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 trial uploads to Babase_test

5b: Return to Excel to troubleshoot errors as needed and save corrections

5c: Assign matured by and ranked by dates and consortdates where appropriate

5d: Censor individuals not seen recently, assign manual census points on exit date where needed, and check for confirmed dispersals

5e: Complete the upload in Babase

Step 6 - Update the BIOGRAPH table with death and censorship information (performed concurrently with Step 5)

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

Also rebuild **(but never empty)** members – this was advisable prior to the addition of residency and now it is necessary to fill in the residency-related rows for newly uploaded data. When making adjustments to a small number of individuals outside of an update it is advisable to just update members for those individuals.

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 to yourself for the next update regarding possible pregnancies
File handwritten sheets in plastic sheath in Demography Update Checklists binder
Make any corrections/clarifications to the Babase 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 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 (if available), the three-month period of the update. Since data now arrives from the field much quicker than it once did, the next month's data will likely not yet be available, which also means that you might need two months of data prior to the start of the quarter to catch any pregnancies that began near the end of the second-to-last month of the quarter. 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 be sure females newly appearing on the sex skin sheets are marked 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); otherwise the tdate is marked with a T. Jeanne marks the end of swelling with a D or Z (**D**eturgescence or conception). The ddate is marked with a Z if the cycle is conceptive and otherwise is marked with a D. 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-conceptive)
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 neonatal binders.

Now that we have switched to quarterly updates and the demographic update for a quarter is usually done before the next month's data even arrives, it is especially important to keep track of those females who may have conceived but have not yet turned P/B by the end of the quarter. It can be easy for the data manager to miss pregnancies that occur near the border between updates (especially if the conception date is in the second to last month of the quarter) if tabs are not kept on the demography update sheets/in the pregs file regarding possible pregnancies.

See Appendix 5 for tips on dealing with cycgaps, Appendix 6 for information about how Jeanne goes about scoring tdates and ddates, and Appendix 7 for assignment of infant birthdates.

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 the 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 the Team to confirm.

1d - Send your list of assigned death dates and dispersal dates around and their associated confidences to Jeanne, Susan, and Beth for approval/discussion.

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 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). See Appendix 2 for the definitions of the various dcauses.

Step 2 - Date Input/Proofing in Excel

2a - Enter "New" Individuals in births 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/c/d) if one doesn't yet exist
3. Copy the last update file (bryya/b/c/d.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 any individuals with census data were conceived and born in the same update period (or both conception and birth are being uploaded this update period because a pregnancy was missed in the last update) those individuals cannot be entered on the summary sheet (see "Important!" below).
7. Once the Team has confirmed all names and sexes of new individuals, consolidate these into the worksheet "Summary". Aborts do not get entrydates or entrytypes while all other new animals start off with entrydate = birth if entrytype is B and entrydate (and statdate) equal to the first day present in a known group in members when entrytype is I or O (you'll have to estimate this and check it later).
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! If a conception and birth for a live individual must be uploaded in the same update period, you will need to upload the conception prior to the birth (which differs from the normal upload order). See the appendix section on ***WHEN BIRTH, DEATH, AND CONCEPTION OCCUR IN THE SAME UPDATE PERIOD FOR FURTHER EXPLANATION.***

Used to Be Important! Babase 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/c/d.xls file for the next update you will do (that way you won't forget to enter them). (Note: You are unlikely to encounter this problem now that we usually don't have the next month's data available so event dates after the last census date for the update will not have been assigned.)

Important! However, it is important to track possible but unconfirmed pregnancies on the demography update sheet to ensure that all are laid out for consideration of whether they are ddates or zdates when preparing for the next update. The last month of the prior

update is always needed to make sure any births that occurred after the last census date can be assigned and to check the cycles that ended in that month to determine whether they were conceptive. Pink on the sex skin is also sometimes slow to appear and be recorded so the prior month is also needed in many cases and once in a while an additional month prior to that.

2b: Enter Census data in a table

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. Create a new census file formatted like the previous one.
2. Copy the column of snames from the last month of the previous update.
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. Do likewise for animals not present but whose exit date will be on a census day.
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 Team is only able to finish an incomplete census (perhaps because the visit was brief or not all subgroups were found). This is indicated on the census datasheet after observer initials and census end time. 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 observers. In these cases the researcher (typically a graduate student) 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 but not for other group members.

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 a given date for the individuals with an N on that date. The presence or absence of these individuals will be interpolated as though observers were 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 Fissions (Protocol used in Babase 1.0 **- no longer in use**)
When a group fissions, 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 and later versions on phpPgAdmin **– now in use**)
Once the group reaches a point where it is more often in subgroups than not, put baboons in the appropriate daughter groups whenever possible and try to keep floaters, especially females, out of group 9 using strategically placed interpolating (M) manual census points. If there is no means of determining which animals were in which subgroups, use the parent grp's gid with an extra 9 on the end for unknown subgroup. Coordinate with the Duke database manager to make sure swerb focal_grp also reflects that the groups are now separate. Continue to enter subgroup notes but also enter intergroup encounters when the groups are together – this way database users can look at the data from the perspective of them all being part of one parent group or from the perspective of them being in separate daughter groups as desired.

Matrilines
Also track changes in group membership on the matrilines in PowerPoint. For each group-year you should have a matriline for use when determining female ranks. You should also have a second matriline for each group for use by the Team (and showing all group members rather than just females). New field matrilines are typically sent to the Team about every six months. They should have enough copies to keep one in the field notebook, one in the office, and one in each vehicle. Matrilines are stored here:
C:\Database Files\ALTMANN\Babase\TEMP\Matrilines\For female ranks
C:\Database Files\ALTMANN\Babase\TEMP\Matrilines\For field use

For ranks matrilines, all members of a given matriline within the group are traced back to their nearest common ancestor. Females who died prior to the year in question but have living descendants are indicated in grey (while females with no living descendants can be deleted in the year after death). Change females who die in the current year to orange to flag them as having died partway through the year (and turn them grey or delete them for the next year's matriline). Add new infants as they are born. For the purposes of ranks we are most interested in females so each female infant should get her own box with her sname and month and year of birth. It can be helpful to know when a male infant has been born or has died so track any last-born male infants in green near the mother's sname – include the month and year of birth but not the sname; include the death mo/yr in orange if the youngster dies. Mark floaters with a purple "FLOATER" below their birth mo/yr.

By contrast in the matrilines for field use, sons get their own boxes within the matriline and males not natal to the group also appear in a box such that all members of the group are accounted for, which means you must also track male movements for the field matrilines. Males' snames and birth dates are indicated in green rather than black. Do not indicate birth years for adult males (including natal males over 8yo, whether they've ranked or not) since the Team annually performs age estimates on adult males but for natal males from other groups do include their natal group and mother's name (Note: annual age estimates are not presently being performed but this matriline rule is still in use). In these matrilines baboons that die or leave the group can simply be deleted.

2c: Enter Matured dates (m and f dates) and Ranked by dates

Fill in the tables 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

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 the last census date is also the statdate in BIOGRAPH). Continue populating the cycles spreadsheets, or if you are creating a new spreadsheet, 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 entry on the Repro Sheet. 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 Demography Update Summary sheets . Parity refers to birth order and usually, but not always, matches the number in the PID. Both the PID and parity should be determined by looking up the female's prior pregnancies to determine what PID and parity are next in the sequence. 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. You can only retrieve the dcpid after you've uploaded the conceptive cycle. Don't enter anything for the resume cycle! Babase will calculate these automatically once a resume cycle has been uploaded. (Babase 1.0 in FoxPro did not do this.)

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

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") (unless the baboon was actually seen in an unidentified 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.

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

1. Field demography notes in each of the study group notebooks
2. Other groups notes in each of the study group notebooks

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 BUT these items are likely worth noting in the index*
• 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, though it is helpful to include group count for good censuses in notes for known individuals*
• Field demography notes providing a verbal description of the field census data (i.e. WAM with Omo's group today or ORI not seen in Linda's group today) without any additional information	*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), although notes that are demographically relevant mainly to one individual (such as the kidnap victim in a kidnapping) are better entered only for that one individual, while the behaviors related to the event should be noted in the index with all participants noted.

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 (10:53)." Include the time if provided in case the animal was sighted more than once that day and so that the entry can be easily matched to any related GPS data.

Other groups notes – Include a record in the Demography Notes Table for any references in the "Other Groups Notes" section of the monitoring notebooks that identify a named baboon, again including the time, if provided, as well as other information of interest such as the group's location or proximity to another group. For example, "GOD was seen in Ositeti group (13:38)."

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 or (3) clearly visited briefly from a nearby group that was not censused that day but in which the individual should be marked present. In some cases, the baboon will not be listed as present in the Members and Census tables in the group visited. For example, "LIB visited Omo's group today (15:46) after being censused in Viola's group this morning." If the visiting baboon sticks around and interacts with group members and there is no other information about his group membership that day, then you can mark him present in the visited group but be sure the baboon is marked absent on the prior and subsequent census dates to prevent Members from interpolating the baboon present erroneously. If observers indicate that the visiting baboon was from another nearby group then the baboon should be marked present in that group.

Floaters moving between subgroups or fission products also may require demography notes of this sort to record their movements between daughter groups over the course of a day.

Behavior data – Behavior data are generally not placed in demography notes unless the behavior has demographic effects, such as kidnapping an infant within or between groups, some other behavior that may contribute to a death, or behavior (such as staying on the periphery of the group or looking settled) that gives an indication of whether a male is planning to disperse or likely to stay. In these cases only the affected individual should receive a demography note (rather than having a repetitive note for every individual involved, as was done for a few years). These events (with all participants listed) and other interesting behaviors should be noted in the demography index (the last tab in the demography notes workbook), along with any other interesting notes about the group as a whole or broader ecological or human trends that may affect the baboons.

New Immigrant Males – Notes on new immigrant males usually have a description of the male including whether he is subadult, a young adult, or an older male and any distinguishing characteristics. Sometimes these notes also indicate whether or not the male is interacting with other group members and in what ways. This information should

be incorporated in a demography note for the male, which should also include the phrase "new immigrant male". Most often these notes only occur on the first and perhaps the second day the new male is present but they occasionally last longer.

Similarly any notes describing visiting females or marker females collared in nonstudy groups should be included in demography notes.

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 its mother, confidence in the death and the date of death, whether the death was witnessed or a corpse found, etc. Thus far all deaths from 2011 onward are in place and older death entries are being backfilled along with the demography note backfill. 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 but if deaths ever becomes a real Babase table, these notes will be redundant and could be removed if desired (though the manual census point placing the animal in the correct group on his or her death date should remain in place).

There is also a table called corpse_info in babase_pending that should be updated with information (collection date, name of animal when known, types of samples collected, whether the skeleton was recovered, notes, skeleton catalog number, and excavation date; as well as, body mass and other measurements thus far only collected for animals who died in Mar 1999) on each corpse or partial corpse found and collected (partly or in its entirety). As of late 2017 there are also now corpse sample lists detailing the samples collected – these might also include useful information about the timing and circumstances of death. Known information concerning collected corpses also needs to be added to Shannon's most recent skeleton collection spreadsheet in Dropbox.

When a baboon with a collar dies the collar_details table in babase_pending also needs to be updated to indicate that the collar has been recovered and to record the recovery date. The active collar Google sheet should also be updated to indicate that the collar is now a spare (if it is reusable) or to remove the collar from the active list.
https://docs.google.com/spreadsheets/d/1dwxKi4kF6lbpYLjn1FgptmKJ2hdd9DA0lUNhzFCbkNQ/edit#gid=0

For efficiency's sake, you will want to enter intergroup encounter notes, usually recorded in the other groups notes section of the group binder but sometimes also appearing in the demography notes, at the same time as you are entering demography notes since the two

datasets are derived from the same notes sections. See the section on Intergroup Encounters for more information.

In March 2016 the Team began entering subgroup membership in the census by assigning the normal X to members of the first subgroup but then the number 2 to members of the second subgroup, 3 to members of the third, etc. This was in hopes that observers would keep more accurate track of which members are in each subgroup than by writing the names of the members in the subgroup notes. With the onset of this change, sleeping subgroup notes have often appeared only in the demography notes (rather than the subgroup notes section) so it often makes sense to enter the subgroup notes at the same time as well. See the sections on Subgroups for more information.

Important! You will need to do a lot of cross referencing, particularly between the demography notes and other sections of the group's data for the month to make sure that anything entered in note form is also noted in the correct place and that all information matches between sources. If everything does not match, ask the Team to clarify. Anything that was not recorded in the correct place needs to be added. Sometimes this will mean simply adding scraps of information to the appropriate entry in another table but sometimes it means creating new entries. You might, for example, need to employ demography notes to create a wounds and pathology entry for a baboon for whom no wounds sheet was ever received. If the out-of-place information is for a dataset handled at Duke, be sure to add the item to the index and highlight it for the Duke database manager so that it can be added to the appropriate table. As of late May 2018 observers are using a new system for collecting interaction data in order to eliminate use of small notebooks by SNS and RSM. Any demography-related notes jotted down on these pages are supposed to be transferred to regular notebook sheets but these should also be checked to be sure everything is present and accurate in the official record.

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 data 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 update 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. Go through all censuses sent for nonstudy groups (including former study groups that have been dropped and any sightings of individuals during other groups collar

checks performed by the drivers) during the update period and highlight all named individuals. Because these groups are not censused with the regularity of the study groups, many individuals are not named or recognized by the 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. It is not necessary to include an actual demography note (i.e., you can leave the notes column blank) unless the animal was seen more than once that day, there is some useful information other than presence included about the baboon (e.g., reproductive information or indication of a wound or pathology), or the group identity is uncertain or needs qualification (e.g., to help differentiate among the various Sinya groups).
   c. Do not include any information about absences unless the quality of the census is good.

3. Likewise enter rows for known baboons the drivers indicate that they *saw* during their Other Groups Collar Checks. If the drivers only heard the individual's collar and did not see him or her then do not create a demography note row.
4. Once you've proofed your entry, put the date in the "Proofed" comment box.
5. Keep an eye out for new infants and female reproductive information for females in recently dropped groups. Infants in dropped groups can sometimes be identified by presence with their mother; this is particularly so for young infants but sometimes older infants can be identified this way in a group that is monitored with some frequency such that we can be certain it is not a newer infant. Assign nontraditional snames (with numbers) to new infants (see the notes in the latest female rank matrilines for naming guidelines) and put any solid reproductive information (swellings greater than 2, P/B, lactating if female has a black infant, etc.) into cycgaps), as well as noting any information about the animals reproductive state or health in the demography notes. Note: estimated conception and birth dates do not go into cycgaps since these were not observed events. However, new mothers in recently dropped study groups should be assigned a demography note on the estimated conception and birth dates to be sure she is interpolated in the correct group on the dates of these events and to allow for inclusion of any notes about the accuracy of the estimates. The infants likewise should receive a demography note on the estimated date of birth, provided they were seen alive at some point. Unless the parity is certain (e.g., female had already conceived before the group was dropped or there is enough data between dropping of the group and the conception to be sure there were no intervening pregnancies), a 100 series pid and parity (like those used for adult females when groups initially came under observation) should be assigned, as the true parity is unknown. For example if a female's last known-parity pregnancy was parity 5, her first uncertain-parity infant would have parity 106.
6. Also include any consortships that were observed in notes for both the male and female participants (if both are named), as well as notes on named males regarding their maturity, ranking, or health since these are not otherwise captured in Babase.

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

Check the baboon project wiki for updates to these directions.

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. Beginning in late 2015 babase_pending is included in the copy. Since Papio was moved onto a departmental server at Duke in Feb 2018 (to comply with security rules), those not at Duke must connect via the Duke VPN (see Duke's VPN website) in order to connect to Papio directly (for commands, *Ranker*, or using R). A Duke NetID is required to connect via the VPN and login to Papio so you will need to obtain a guest NetID.

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 YOURADMINNAME
babase_test babase

(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 minutes or so, but it is working.

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, ranked by dates to RANKDATES

9 – Upload dispersals to DISPERSEDATES, first consortships to CONSORTDATES (tables)

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 and MEMBERS to calculate residency (but rebuild MEMBERS only in real Babase so you don't crash the server)

15 – Upload manual weather data to MIN_MAXS

16 – Upload WeatherHawk files to WEATHERHAWK

17 – Upload wounds and pathologies files to WP_REPORTS_OBSERVERS, WP_DETAILS_AFFECTEDPARTS, WP_HEALS (views)

18 – Upload neonatals to NEONATALS in babase_pending

19 – Upload subgroup notes to SUBGROUPS in babase_pending

20 – Upload intergroup encounters to INTERGROUP_ENCOUNTERS in babase_pending

21 – Upload raw hybridity scores to HYBRIDITY_RAWMORPHO in babase_pending

22 – Empty the DEATHS table in babase_pending and re-upload all deaths (since rows for older deaths are still being filled in, this ensures that all available information is present in the babase_pending table)

23 – Upload sex skin data to SEXSKINS and PCSKINS (if any are ready)

24 – 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 sometimes, especially regarding animals whose bodies are recovered, from notes, emails, or corpse sample lists – introduced in Sep 2017 to aid in tracking of samples collected from corpses) and used to update BIOGRAPH. Unlike prior steps, you do not prepare a "deaths" table to upload (well, okay, now you do upload rows to a babase_pending table but that's not how you actually "kill" the baboons in the database…). This is one of the few cases where standard protocol is to manually edit the data in the statdate, status, dcause, dcausenatureconfidence, and dcauseagentconfidence 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

Dcausenatureconfidence and decauseagentconfidence – 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, 3, or 4. New data can be entered after the statdate for individuals considered to be alive (status =

0), in which case 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.

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 (…or fate unknown for the rare status 4 individual). 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 as 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; thus, we prevent data from accidentally being uploaded for individuals after they are no longer part of the study population. Three study groups (Kelly's, Mica's, and Snap's) were dropped in 2012 and Dibble's group was dropped in 2019. Members of these groups are also being censored as they drop off the other group census sheets (i.e., as the field team loses IDs for these individuals). 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 also sometimes happens with females in dropped study groups.) 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 if the exit date was assigned by picking a midpoint between the last date present and the first date absent so that each baboon is present in the correct group on its dispersedate or final exit date. Use interpolating (M) points for deaths and noninterpolating (N) points for dispersals and

censorships. 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 also 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 statdate from date last seen alive) during demography note backfill, which also includes assignment of dcausenatureconfidence and dcauseagentconfidence.

Criteria for declaring an adult male or male of at least early dispersal age (4yo) who is not seen again are as follows:

1. The male's remains and/or collar were recovered (e.g., Elvis, Fabian, and Lofty).
2. The male was seriously injured or ill or very old and in decline when last seen (e.g., Khan, Powell, Bock, and Alex).
3. Male was a stable resident of the group who disappeared along with other individuals known or suspected to have died (e.g., Beam, Cool, and Wegner).

Otherwise males of at least four years of age who disappear and are not seen again are assumed to have dispersed and are censored.

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.
But in 2017 we changed our minds about that because occasionally fishy
things were happening that affected entrydates so now we're rebuilding
members again. And then at the end of 2019 we added residency to
members and a rebuild is required to update the residency columns.
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.

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 ranks 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).

From time to time females who aren't regular group members may show up in the dominance matrices. If the female was not ranked in the last month's ranking, she will show up at the bottom with the new infants so a careful eye must be kept for unexpected individuals with brief or occasional membership in the group. This mainly occurs during and soon after a fission when some females (often young females but sometimes more mature members as well) may float between groups for a time or make a short visit to the fission product they're not usually with. More rarely we may also have a young or socially isolated female make a short visit to another group or groups, sometimes accompanying a male from her group.

With individuals floating between fission products, there should be some basis for ranking the female in her nonstandard group since the members of the two groups are sometimes or were recently together. Thus floating individuals should be ranked in whichever daughter groups in which they appear.

In the case of females taking odd trips to another group, however, there may be little-to-no basis for ranking the visitor so these females are often excluded from ranks (e.g., Ceejay visiting Linda's group or Dibble's periodic visits to other groups). The following rules were established during a leaders' call and disseminated via email on 2 Jun 2016:

Female visitors

1. will not be ranked for a month in the group being visited if the visit is less than two weeks. Consequently, Dibble will not be ranked during her visits to Vogue's in July 2015 or Narasha's in January or February 2016.

2. will potentially be ranked in a month if the visit is longer, but this will be considered on a case by case basis, considering what interaction data are available and whether, for example, the visitor's home group and the visiting group were fairly recent fission products of the same group. We will review these cases as they occur and consider changes to the rules based on our experience if these cases become more common.

To leave a female out of ranks, simply place her at the bottom of the ranks in *Ranker* and then delete her row from ranks.

The following 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 but please see Decision rules for assigning ranks below.

1. Make sure that both focal and decided agonism data have been updated through the rank time period (and therefore the demography update will also have been completed).
2. As you input the census data for each group-month, you should have updated the female rank matrilines for that year in PowerPoint here: C:\Database Files\ALTMANN\Babase\TEMP\Matrilines\For female ranks (as well as those for use by the field team located in the neighboring folder called For field use). This will provide extra context (e.g., age and identities of maternal relatives) 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 and to review any rank decisions that were uncertain due to a lack of forward-looking data.
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, particularly in tumultuous years. 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 to Babase as you work through the months – *only save to text*. 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.

Ranks During Fissions

Until the Nyayo's fission, census and ranks were starkly divided between parent and daughter groups with the fission magically happening on a particular date. But with this recent round of fissions (and this is gradually being backfilled to older fissions), once a fission is reasonably under way, the membership is divided between parent and daughter groups depending on how the group was found each day. This means that the parent and daughter groups overlap in time and individuals may be present in more both parent and one or more daughter groups within a given month. Thus during a fission it is typically necessary to have ranks for both the parent and the daughter groups within the same month for one or more months while the fission is in progress. Male and female ranks should be coordinated so that they are done for the same group-months.

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 and (strongly preferred) the first quarter of the following year.

Using *Ranker*

Step 1: Connect to the Duke VPN
Step 2: Create an SSH tunnel to Papio
Step 3: Launch *Ranker*
Step 4: Create 16 ALF Matrices for each group
Step 5: Manipulate the rank order
Step 6: 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

History and Overview

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* was the first version of the *Ranker* program to be integrated with Babase 2.0, available online on phpPgAdmin. It was developed at Duke by undergraduate Tylor Brock and updated by his advisor, Jun Yang, and later by Karl Pinc and Jake Gordon. Contact the Duke database manager for the current version of Ranker.

At Princeton, we do All Female Ranks (ALF) and Adult Female Ranks (ADF). Duke is responsible for the male ranks (ALM). A description of the male rank protocol should be available on the Wiki in the Duke data management protocol.

Always create the ALF ranks first. In Babase 1.0 there was also a program called "SubRanker" that extracted the ADF from the ALF ranks. This is not a separate program in *Ranker* 2.0 and beyond. Rather, you just select ADF from the dropdown menu at the beginning to seed the program for this – you can then seed the ADF ranks with the ALF ranks.

The database manager creates the agonism matrix outputs (using *Ranker*) and does a first pass on rank changes. The suggested rank changes then go to Jeanne for review. She may 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. For a difficult group-year there could be several versions of the agonism matrices before Jeanne settles on a 'final' ranking for each month and some dyadic pairs may be marked for review when the following year's ranks are assigned. 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*

The female rank assignments are done by year. *Ranker 2.0* and later iterations create interaction matrices by drawing agonism data directly from Babase. Therefore, all agonism data for the entire calendar year must first be uploaded to Babase. (This is part of Duke's quarterly 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 Duke *VPN* program and create an SSH tunnel (both using a Duke NetID) to access and run *Ranker*. The protocols to for this are available on the Babase Wiki. If you don't have a Duke NetID, contact your friendly neighborhood Duke PI to sponsor you for a NetID.

Step 1: Connect to the Duke VPN

In the bottom toolbar right click on the Duke VPN logo and select "Connect". These dialogue boxes will popup.

Type in your Duke NetID and password and hit OK. Throughout this process Duke will have warnings popping up about needing to be an authorized user, etc. If it's in a pop up box, close it to acknowledge that you have received the warning so you can move on.

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.

Step 2: Open SSH Tunnel

Create an SSH tunnel using the command below (for Windows users, this requires an SSH client – see the wiki if you need directions on setting one up – login using your Duke NetID).

ssh -L 5432:localhost:5432 NetID@papio.biology.duke.edu

Step 3: Launch *Ranker*

Click on the LaunchRanker file written by Jesse Saunders on the database manager's desktop. The *Ranker* window should then open.

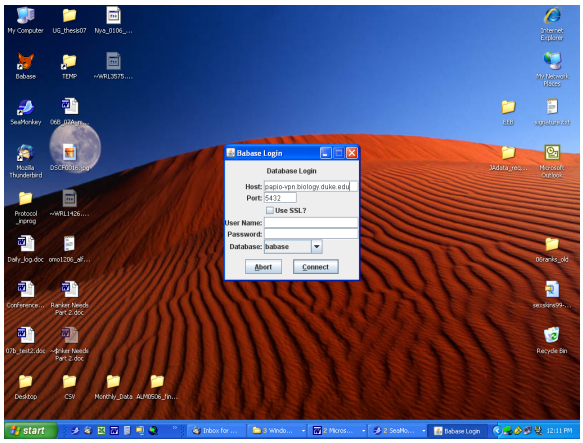
.

Don't close any of the SSH client windows or terminals, as this will disconnect you.

Select

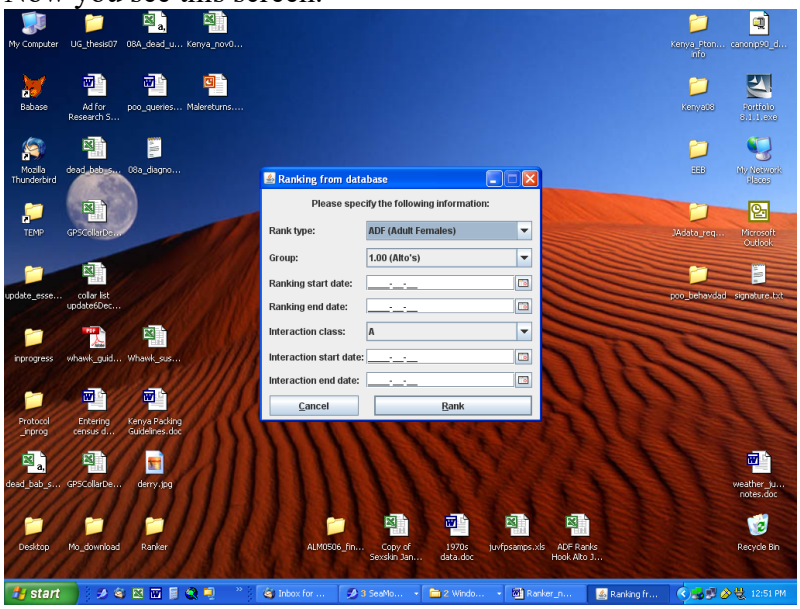
Create/load ranking from Babase

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.



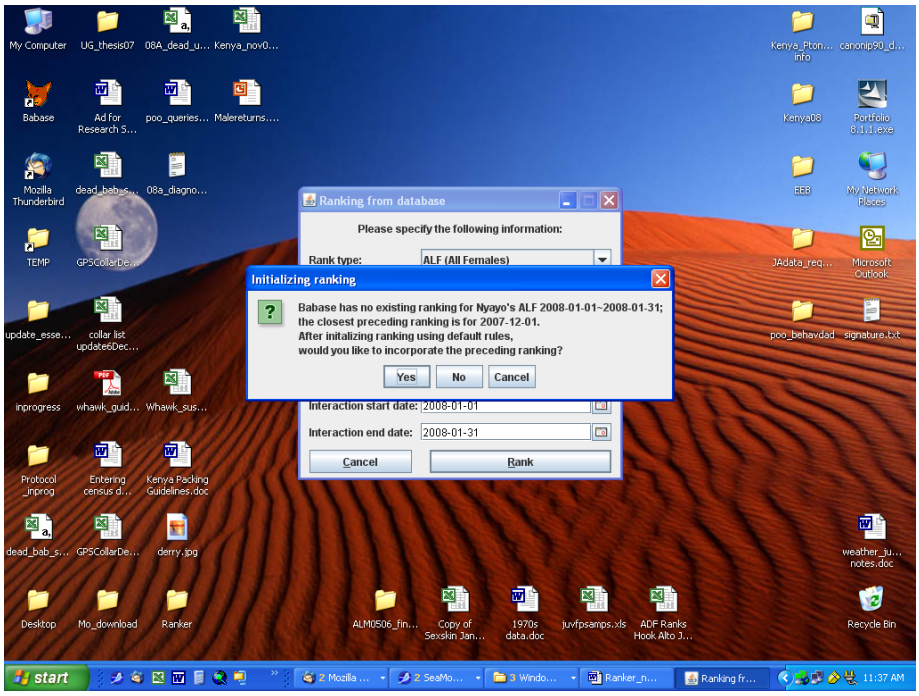
Step 4: Create ALF Matrices for Each Group

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.

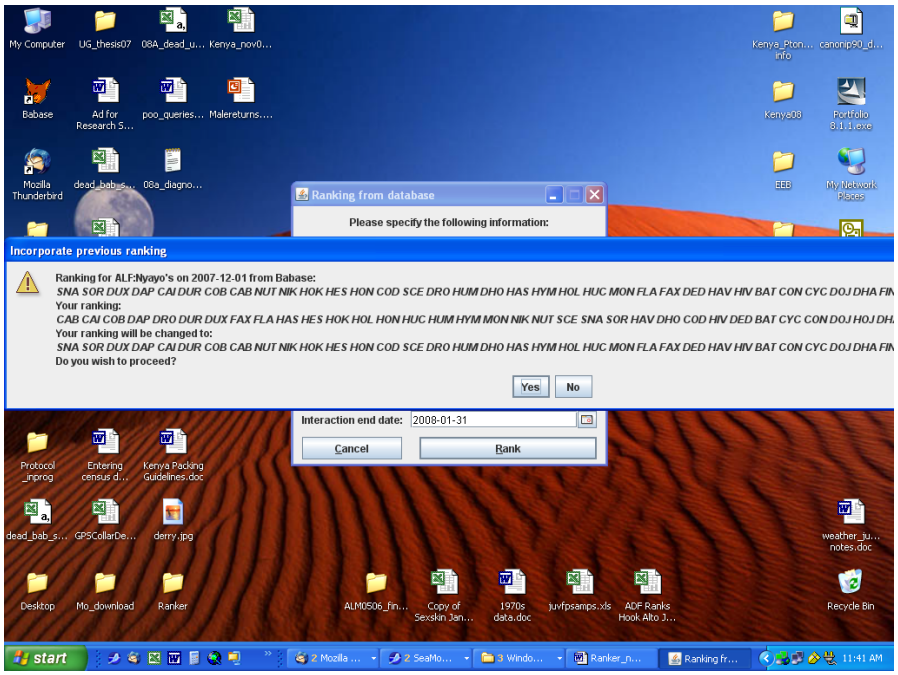
- Choose the ALF option in the first box.
- Select the group to be ranked in the second box.
- Put the range of dates of interest in the 3rd and 4th boxes (if this is your first matrix, do January (e.g.,. start date 2008-01-01, end date 2008-01-31)

- Leave the default interaction class as ‘A’
- Repeat this range of dates in the 5th and 6th boxes.
- Hit Rank

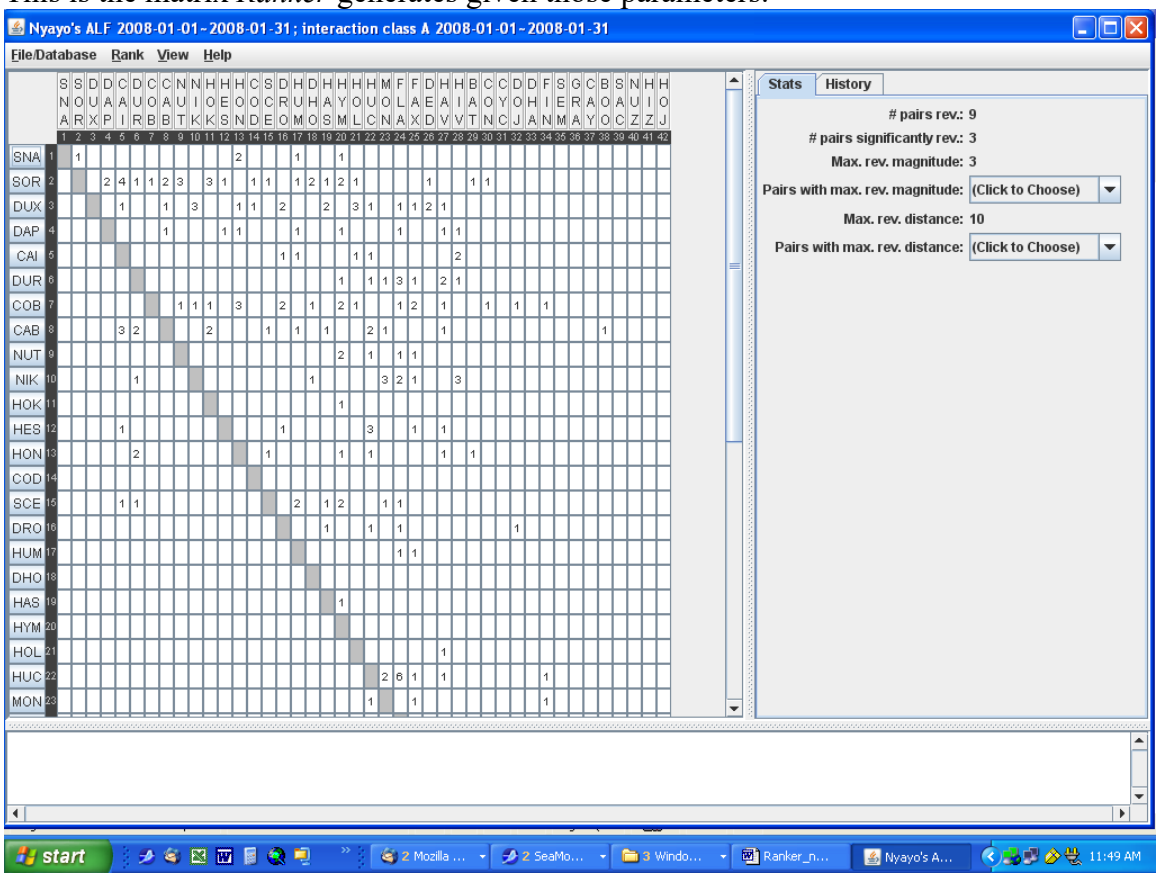
Because there is no existing rank in Babase for January of 2008, you will get a message like the following. Answer YES, you want to use the rank from the previous month. (You should have a December rank order. If it’s missing, you never input it from the last ranks update.)



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* 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 Decision Rules for Assigning Ranks.

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 5: 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) you 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 6: 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 adjust the font size downward and/or adjust margins as needed for larger groups. You need to see the entire matrix on a single page (though 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.

CONVENTIONS FOR NAMING MATRICES

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. Older matrix files are named thus:

Example: a temporary matrix for Weaver's Jan 2006 agonism is **wt0106**

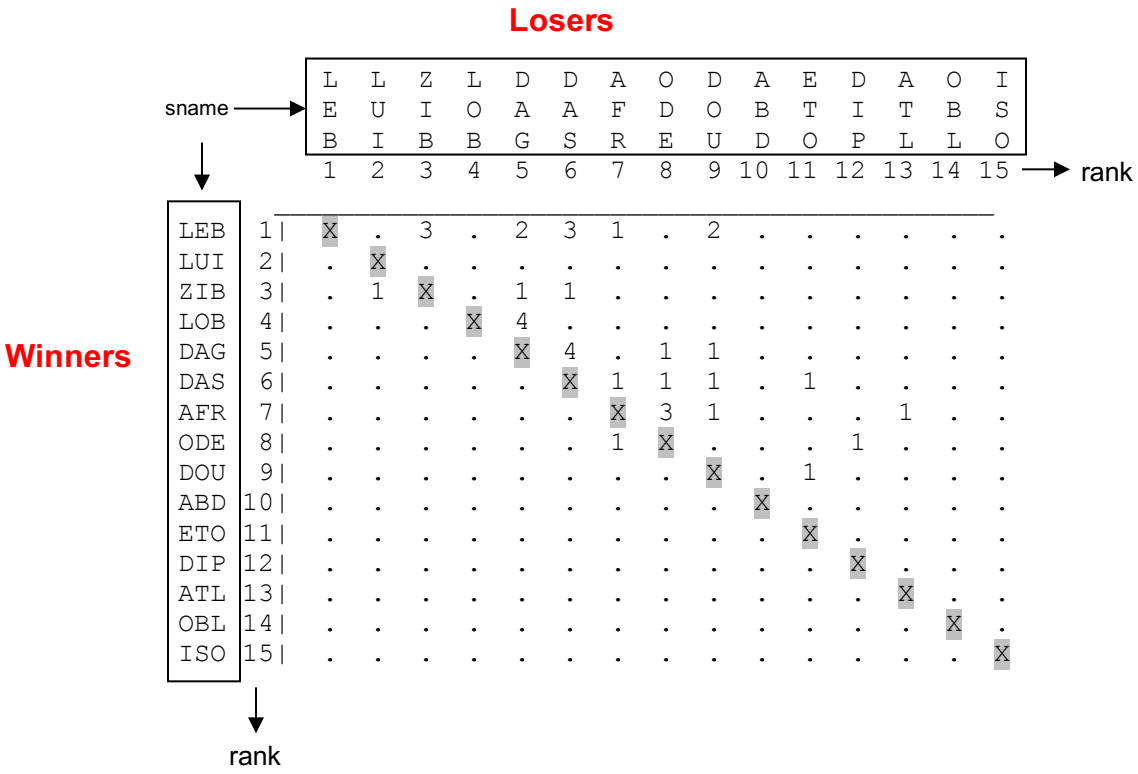
The convention for the cumulative matrices is a little more complicated
wt010606 – covers January (**01**) to June (**06**) of 2006 (**06**)
wt071206 – covers July (**07**) to December (**12**) of 2006 (**06**)
wt011206 – covers January (**01**) to December (**12**) of 2006 (**06**)

This naming system became somewhat cumbersome after quarterly updates were introduced since the process of assigning female ranks is now often begun once the first quarter of the following year's agonisms are available in Babase. Thus the naming convention was updated as follows:

t01_2014_Aca = temporary January 2014 matrix for Acacia's group
t2014b_Vog = matrix for July through December for Vogue's group
2015_Q1_Nar = matrix showing the agonisms for Narasha's in 2015's first quarter
f2014_FY_Hok = final full year matrix for Hokey's group
r2013_12_Dib = revised matrix for Dec 2013 for Dibble's group (revised when 2014 ranks were completed)

READING THE RANK MATRICES

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.

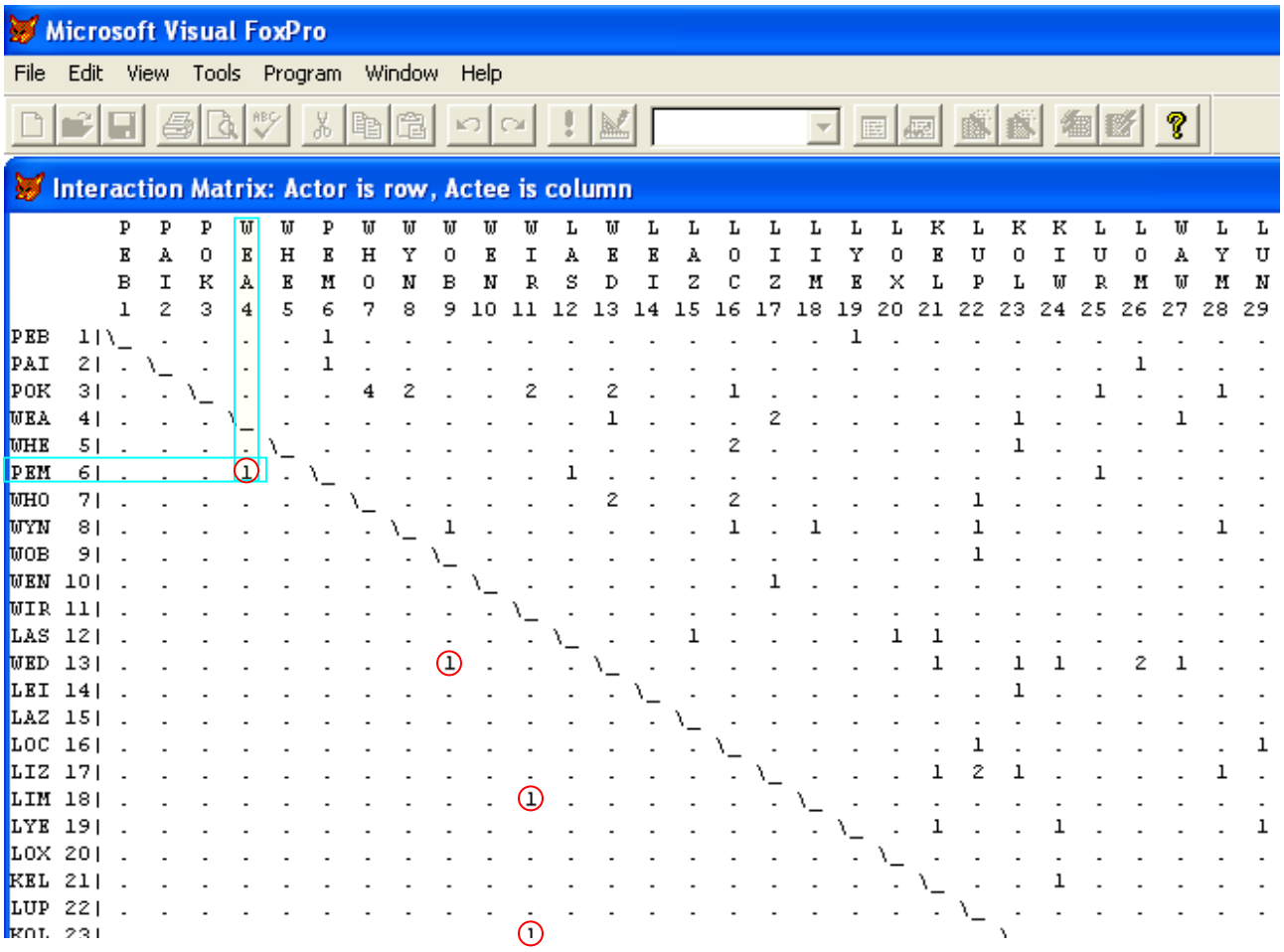


		LEB	LUI	ZIB	LOB	DAG	DAS	AFR	ODE	DOU	ABD	ETO	DIP	ATL	OBL	ISO
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LEB	1	X	.	3	.	2	3	1	.	2	.	.	.	.	.	.
LUI	2	.	X	.	.	.	.	.	.	.	.	.	.	.	.	.
ZIB	3	.	1	X	.	1	1	.	.	.	.	.	.	.	.	.
LOB	4	.	.	.	X	4	.	.	.	.	.	.	.	.	.	.
DAG	5	.	.	.	.	X	4	.	1	1	.	.	.	.	.	.
DAS	6	.	.	.	.	.	X	1	1	1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

IDENTIFYING ALL RANK REVERSALS

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.

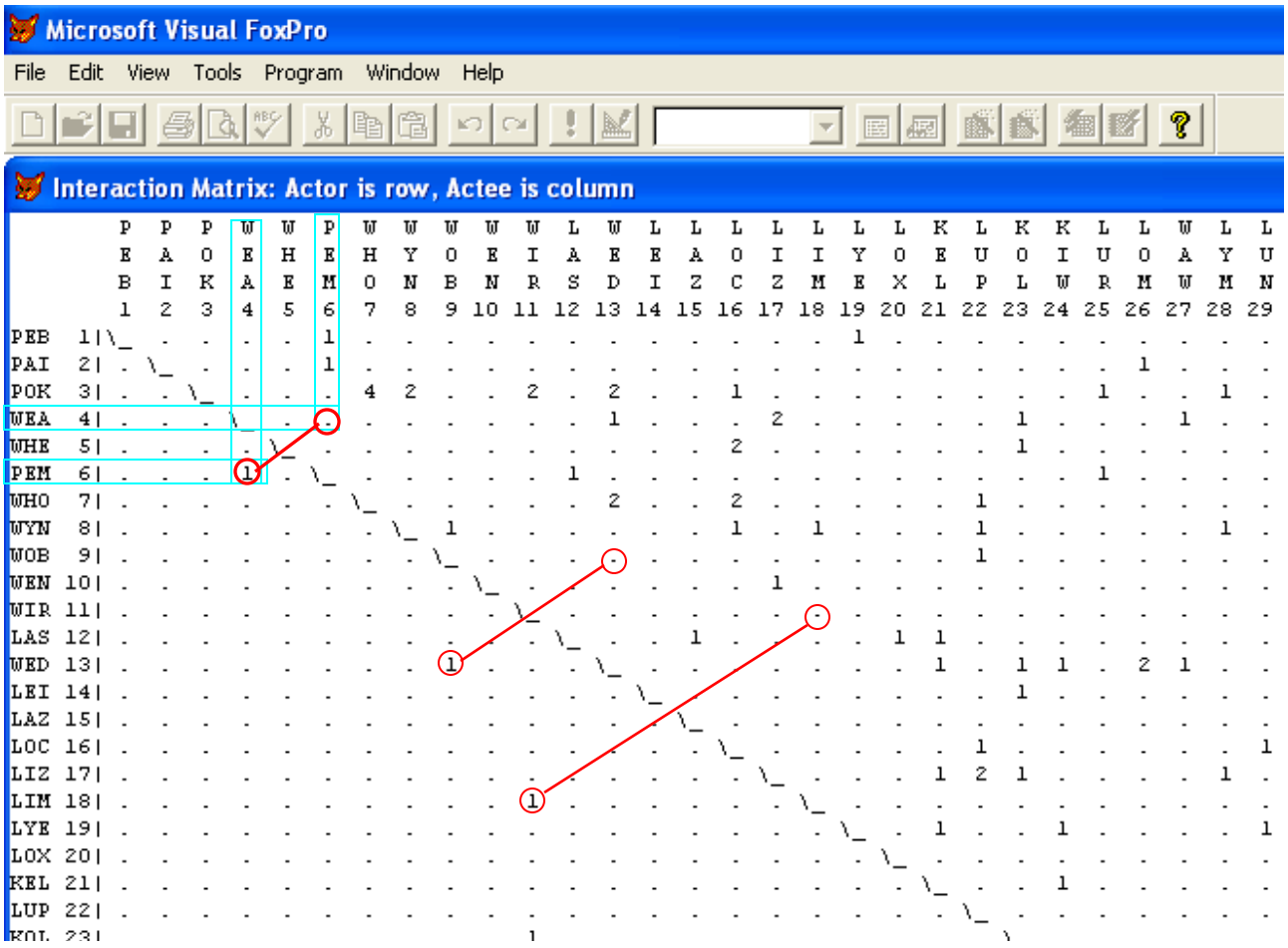


		PEB 1	PAI 2	POK 3	WEA 4	WHE 5	PEM 6	WHO 7	WYN 8	WOB 9	WEN 10	WIR 11	LAS 12	WED 13	LEI 14	LAZ 15	LOC 16	LIZ 17	LIM 18	LYE 19	LOX 20	KEL 21	LUP 22	KOL 23
PEB	1	_	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
PAI	2	.	_	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
POK	3	.	.	_	.	.	.	4	2	.	.	2	.	2	.	.	1	.	.	.	.	.	.	.	1
WEA	4	.	.	.	_	.	.	.	.	.	.	.	.	1	.	.	.	2	.	.	.	.	1	.	.
WHE	5	.	.	.	.	_	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	1	.	.
PEM	6	.	.	.	1	.	_	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.
WHO	7	.	.	.	.	.	.	_	.	.	.	.	.	2	.	.	2	.	.	.	.	.	1	.	.
WYN	8	.	.	.	.	.	.	.	_	1	.	.	.	.	.	.	1	.	1	.	.	.	1	.	.
WOB	9	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
WEN	10	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	1	.	.	.	.	.	.	.	.
WIR	11	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	.	.	.	.	.	.	.
LAS	12	.	.	.	.	.	.	.	.	.	.	.	_	.	.	1	.	.	.	.	1	1	.	.	.
WED	13	.	.	.	.	.	.	.	.	1	.	.	.	_	.	.	.	.	.	.	.	1	.	1	1
LEI	14	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	.	.	1	.
LAZ	15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	.	.	.
LOC	16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	1	.	.
LIZ	17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.	1	2	1	.
LIM	18	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	_	.	.	.	.	.	.
LYE	19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	1	.	.	1
LOX	20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.
KEL	21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.
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.



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.

DECISION RULES FOR ASSIGNING RANKS

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?
- Does credit cause new reversals with the baboons in between?
- What other wins or losses do the involved individuals have?
- 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)?
- Has something happened that may have temporarily destabilized the hierarchy?

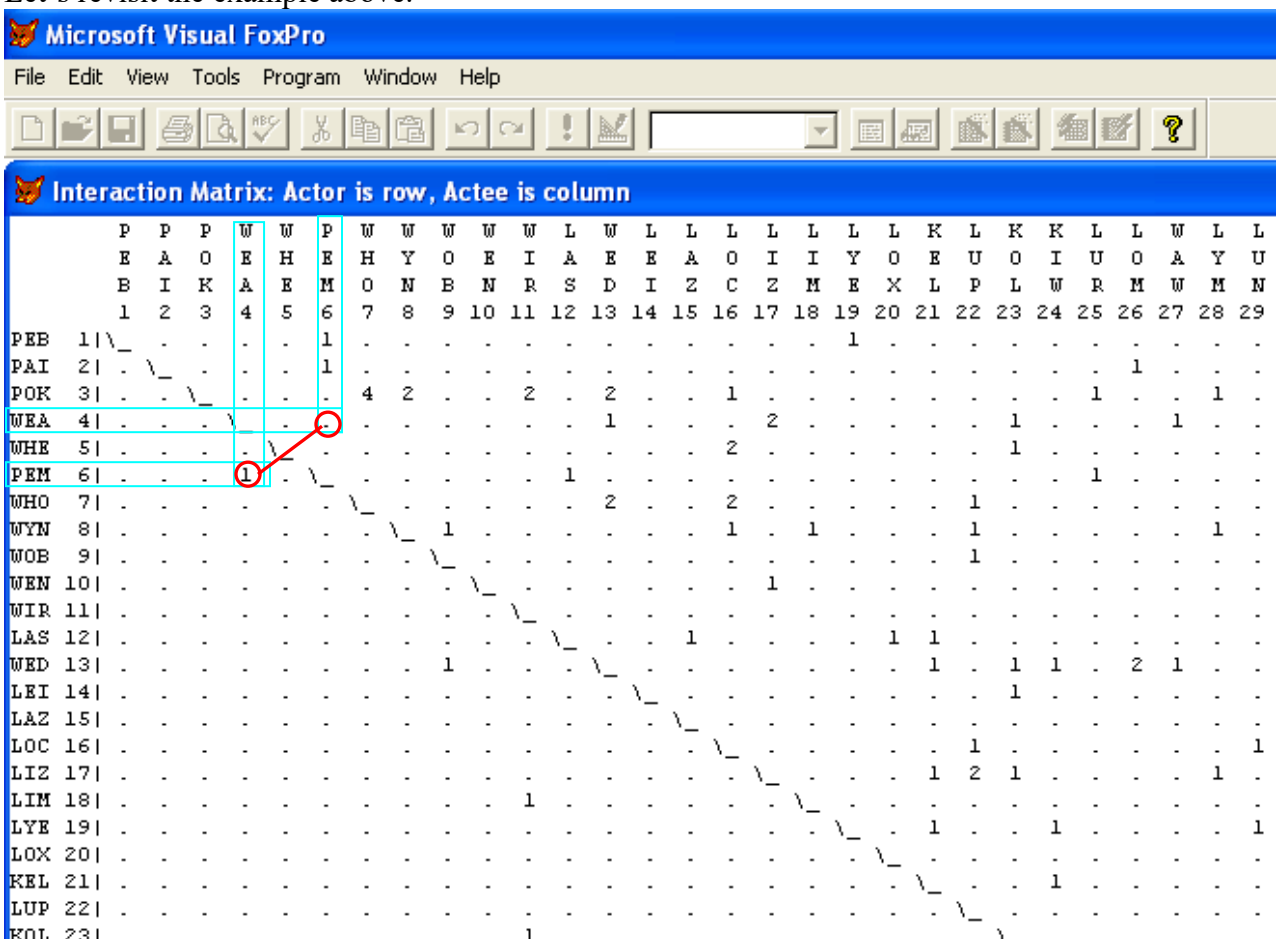
When there is a need to look forward and the agonism data are not yet in Babase, review the monthly matrices that are sent from the Team in Amboseli to look for pertinent trends.

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.

A Sample Reversal

Let's revisit the example above.



		PEB	PAI	POK	WEA	WHE	PEM	WHO	WYN	WOB	WEN	WIR	LAS	WED	LEI	LAZ	LOC	LIZ	LIM	LYE	LOX	KEL	LUP
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
PEB	1	_	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
PAI	2	.	_	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
POK	3	.	.	_	.	.	.	4	2	.	.	2	.	2	.	.	1	.	.	.	.	.	.
WEA	4	.	.	.	_	.	.	.	.	.	.	.	.	1	.	.	.	2	.	.	.	.	.
WHE	5	.	.	.	.	_	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.
PEM	6	.	.	.	1	.	_	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
WHO	7	.	.	.	.	.	.	_	.	.	.	.	.	2	.	.	2	.	.	.	.	.	1
WYN	8	.	.	.	.	.	.	.	_	1	.	.	.	.	.	.	1	.	1	.	.	.	1
WOB	9	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	.	.	.	.	.	.	1
WEN	10	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	1	.	.	.	.	.
WIR	11	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	.	.	.	.	.
LAS	12	.	.	.	.	.	.	.	.	.	.	.	_	.	.	1	.	.	.	.	1	1	.
WED	13	.	.	.	.	.	.	.	.	1	.	.	.	_	.	.	.	.	.	.	.	1	.
LEI	14	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	.	1
LAZ	15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	.	.
LOC	16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	.	.	.	1
LIZ	17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.	1	2	1
LIM	18	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	_	.	.	.	.
LYE	19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	1	.
LOX	20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.	.
KEL	21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_	.
LUP	22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	_
KOL	23											1											

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 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 (if needed) 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.

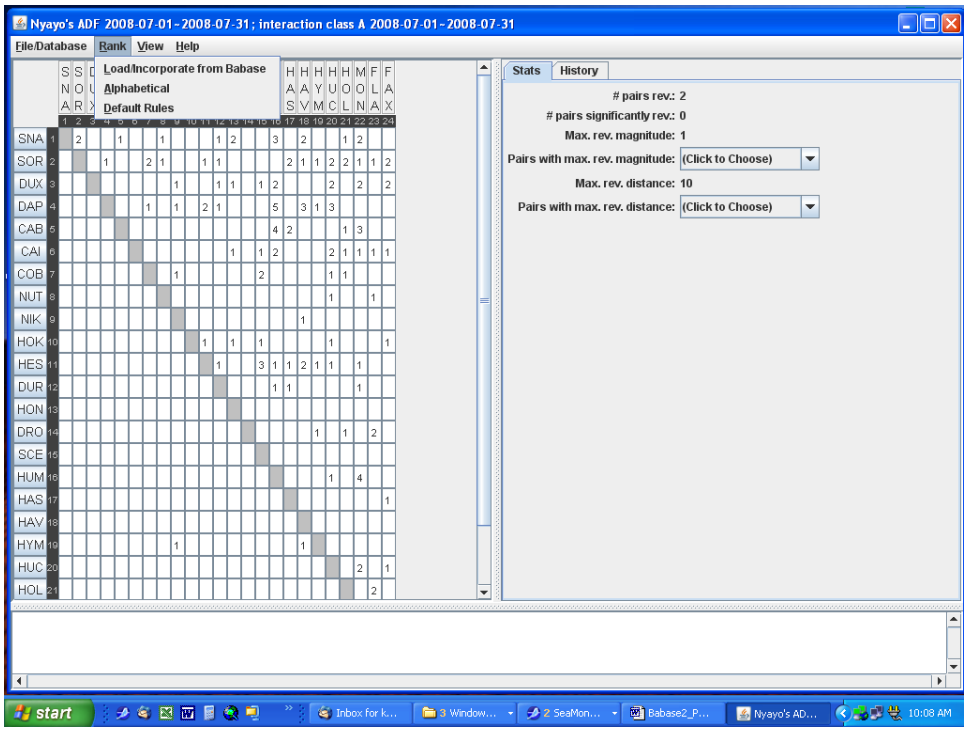
HELPFUL HINT
Once you decide a rank change should occur, look at the cumulative matrix and go through the stickies for all the reversals in each month until the end of the year. If your decision has solved a problem or changed a reversal that exists in later months, mark the relevant reversal in the later month as "Credit already given" and indicate when. 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.

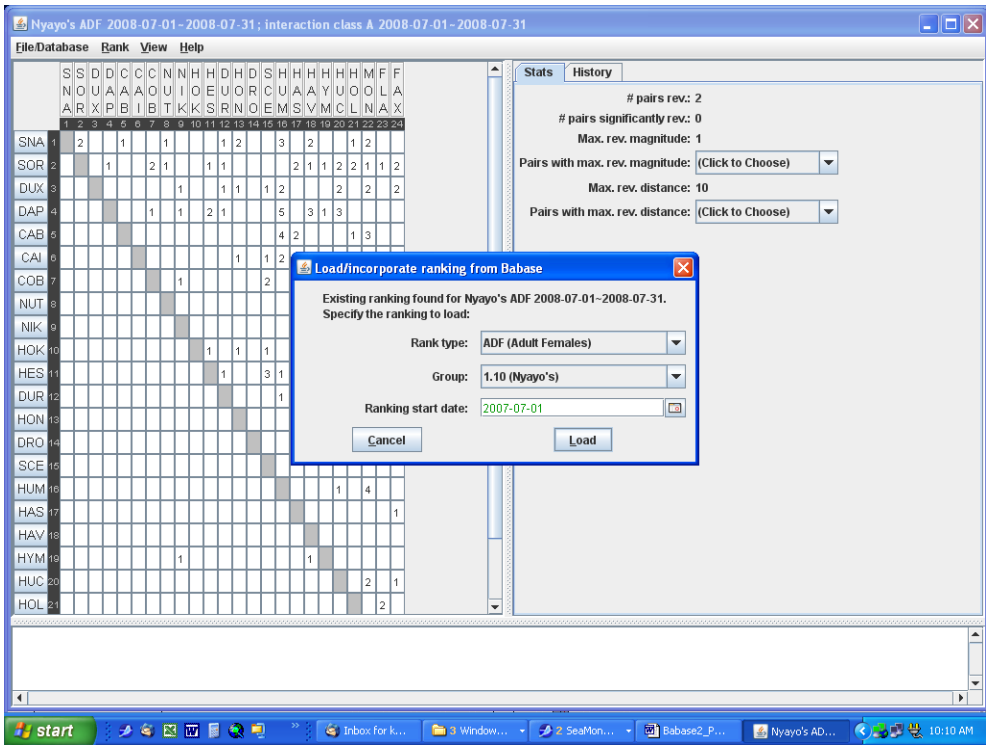
MANIPULATING THE RANK ORDER (UPDATED 2009-06-18)

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

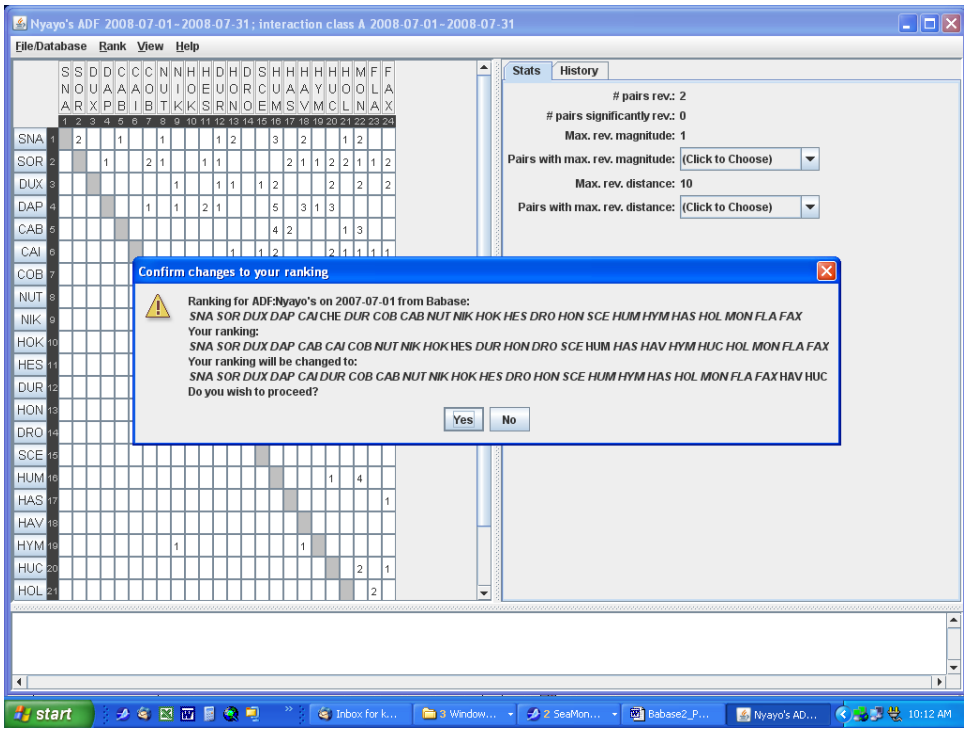


In the dialogue box select the options you want from the drop-down boxes. In the example below, I'm going to examine how the agonisms from July 2008 would look if I used the rank order from July 2007.



Click Load

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



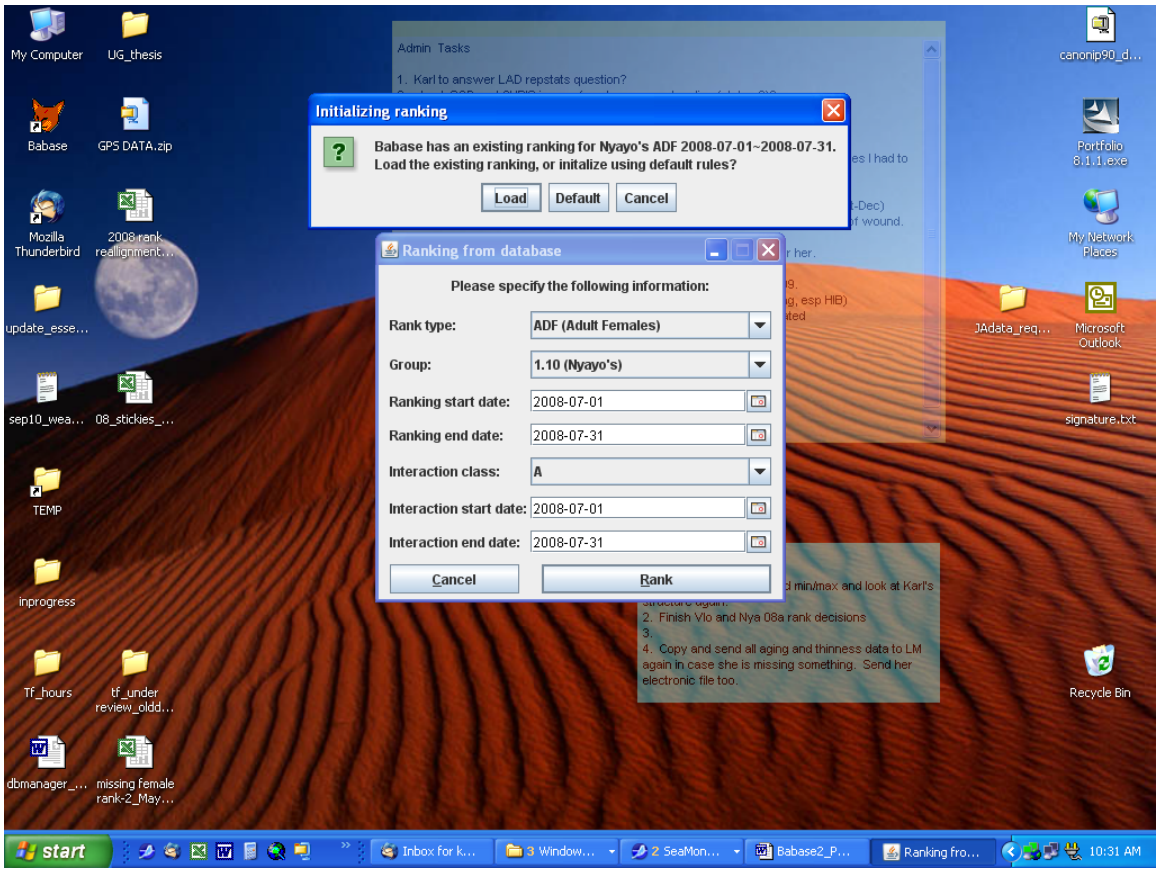
Ranker allows you to also do manual changes of an individual's rank in order to explore how wins/losses change within the matrix. This is useful in the second and third versions of the matrices as you begin manipulations of rank order. Drag and drop snames to move individuals up or down the ranks.

RANKING THE ADFS (UPDATED 2009-06-18)

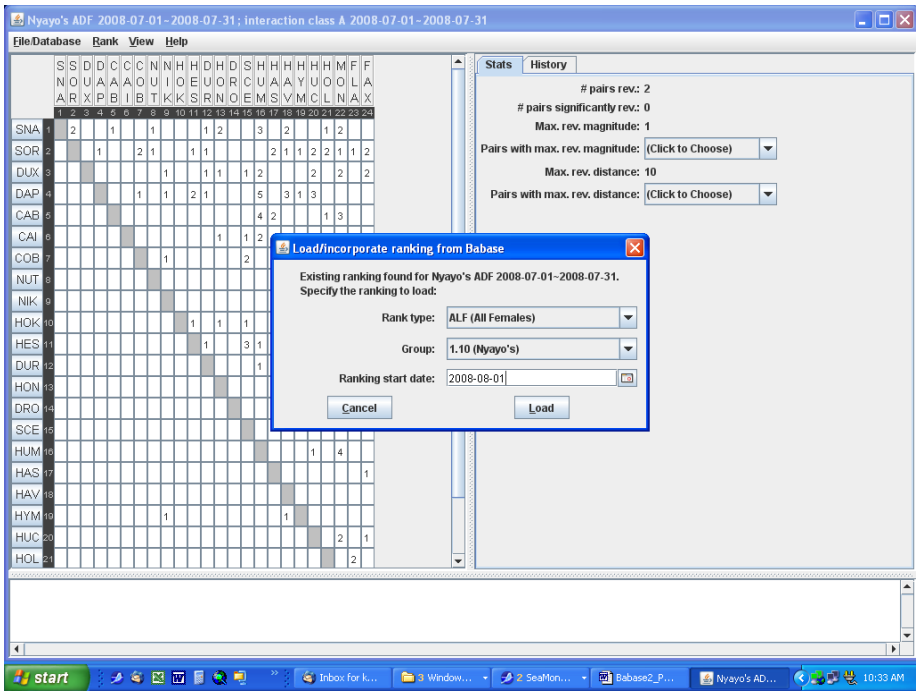
Female ranks are entered and updated in Babase using the *Ranker* programs for all female ranks (ALF) and adult female ranks (ADF), respectively. The *Ranker* program displays decided agonism data for a particular group and time period in a win/loss matrix showing the number of times every individual 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

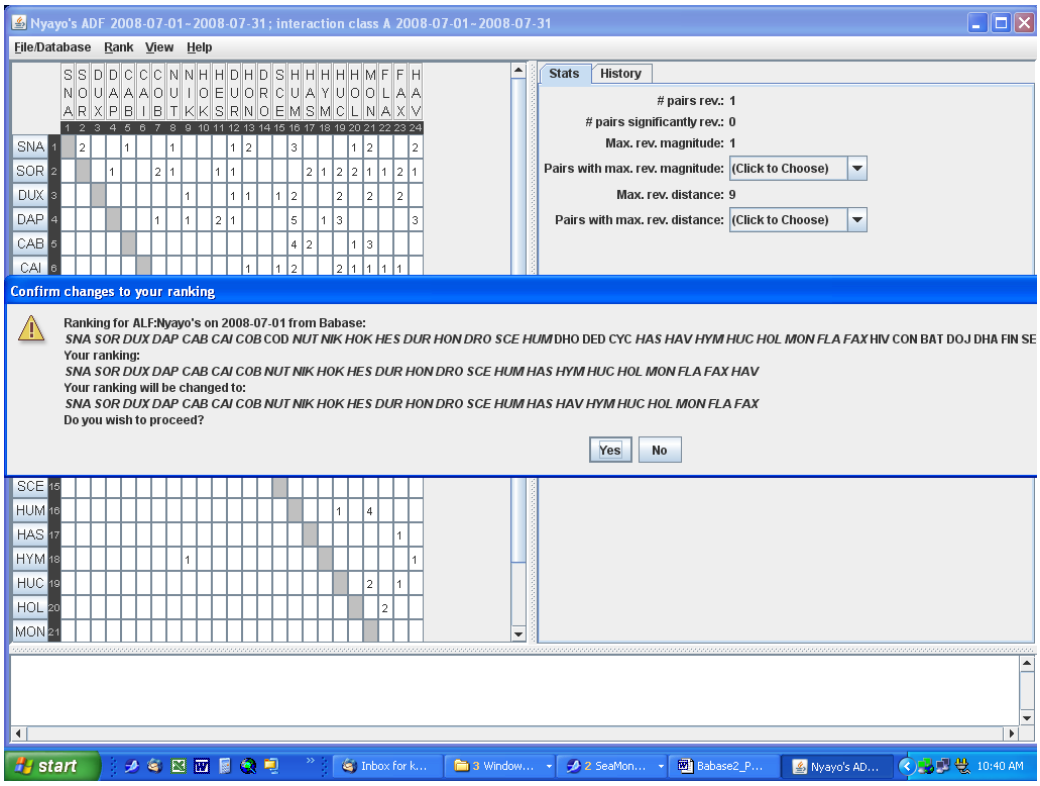
To do the ADF rankings, create a matrix for your month of interest using the ADF option:



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 three 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 previously that was never integrated into Babase 1.0 in FoxPro; however, dates with no readings are not entered into Babase 2.0 and later versions.

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 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.

METEOROLOGICAL DATA

WeatherHawk Weather Station
The WeatherHawk Weather Station automatically takes hourly readings on a number of meteorological measurements. The Team in Kenya used to download this information into daily CSV files. These files, in turn, were 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, as the new software allows the Team to send one file weekly. These weekly files are saved to the monthly field data folder and appended to the master WeatherHawk file monthly for integrity checking. The data are uploaded to Babase quarterly. Additionally, receipt of the data is confirmed during the Babase call on a weekly or semi-weekly basis. The below description includes the old process and the changes over time.

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 text 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 or the file is in Dropbox but not emailed. 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.

OLD SOFTWARE (VIRTUAL WEATHER STATION) 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

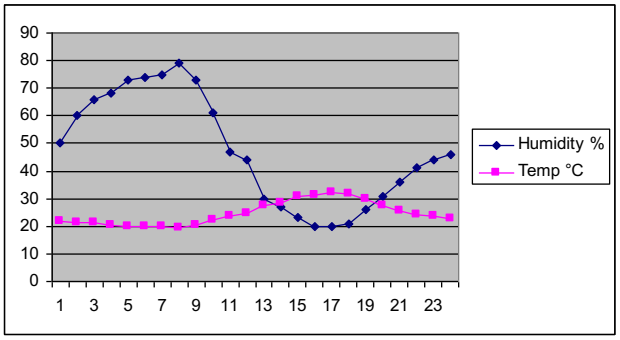
RecNum		Reference number for the data row
BatVolt_V	volts	Battery voltage (hourly)
BatVolt_V_Min	volts	Minimum hourly battery voltage
AirTemp_C_Avg	°C	Average hourly air temperature
RH_Avg	%	Average hourly percent humidity
WindSpeed_ms_Avg	m/s	Average hourly wind speed
Solar_Avg	W/m^2	Average hourly 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 wind speed
Barometer_KPa	kPa	Atmospheric pressure
RainYearly_mm	mm	Cumulative annual rainfall (in 1mm increments)

Some problems with sensors have occurred in the past, as well as one battery failure. To track sensor changes and failures we have a table in babase_pending called WEATHERHAWK_SENSOR_CHANGES_AND_FAILURES, which should be updated whenever JKW alerts you that he has changed a sensor or battery, or when a failure is discovered. The temp/rh sensor is shelf stable and needs changing the most frequently (annually) so there is typically a spare in camp with a new one sent each year, while the barometer and battery should be sent as needed according to their replacement schedules.

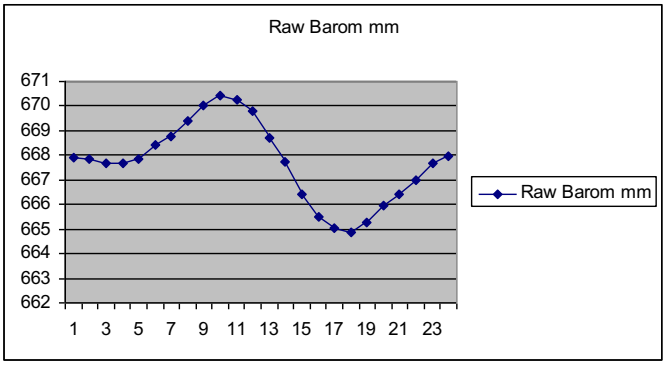
You can do a quick and dirty visual check on most of the raw data columns by making some simple graphs.

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

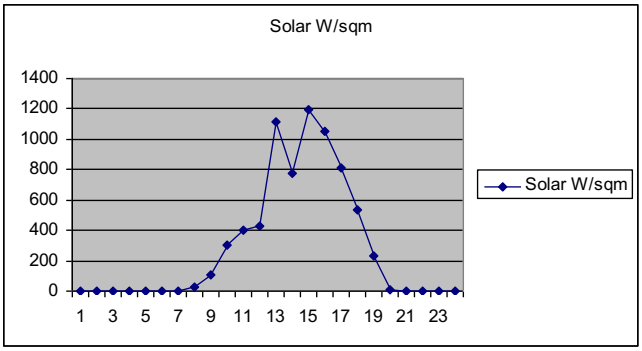
than absolute value. Expect higher humidity in the wet months and lower in the dry months.



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.



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.



Also keep an eye on dips in the battery voltage. Since it is solar powered, it follows a regular pattern of dipping a little at night, but if it begins to dip further, that is a sign of impending battery failure and a new battery should be sent at the earliest opportunity. The battery should last for at least five years but we have had one fail early, resulting in a loss of overnight and early morning data (with longer gaps if the morning was cloudy) for several months before it was discovered. But a spare battery cannot be kept in camp because it does not last long in storage. Both the BatVolt and BatVolt_Min columns may be informative. To help raise the alarm, a warning was added to integrity_warnings (see Warning System) that throws warnings whenever BatVolt drops below 12.5 or BatVolt_Min drops below 12.45.

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.

WeatherHawk data resided for many years in babase_pending (where there were two versions, one containing all data collected using the old software and one containing 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) but were moved to weatherhawk in Babase in Mar 2017. This also necessitated creation of the support table weatherhawk_softwares to indicate for each row which software was used to collect the data.

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 had to be resolved in some manner when WeatherHawk data was 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.

We decided that only wind speed was too problematic to simply convert data from the old software to the units of the new software. Thus, there are two columns for wind speed, one in km/h for data from the old software and one in m/s for data from the new software. For all other columns included in the weatherhawk table, units were converted to those used by the new software.

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: The above-mentioned 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.

Weatherhawk data are uploaded during the quarterly demography update.

There is also now a warning in the Warning System to increase the odds of early detection of impending battery failure by producing rows when the voltage begins to drop lower overnight. Occasionally this happens with a healthy battery during very cloudy periods since the battery cannot charge as well during the daylight hours. The batteries are meant to last 4-5 years but are not guaranteed. One began to fail at only about 3.5 years in 2017, which is when we added the warning. It successfully detected a decline in battery performance of the replacement battery only 3 years later.

WEATHERHAWKS_SENSOR_CHANGES_AND_FAILURES – Be sure to update the WEATHERHAWKS_SENSOR_CHANGES_AND_FAILURES table in babase_pending at least annually. JKW should send an email noting that he has changed the temperature/pH sensor and any other parts due for a change. Inquire if no word has been received and a change is overdue. Also update it whenever data are lost for a given sensor or the whole device for any measurable period.

MALE AGE ESTIMATES

Male age estimates, both age estimates for new immigrant males and annual age estimates on resident males (conducted for many years, usually 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. The annual assessments were first done in 2005 and then in 2007 through 2014. They have not been performed since.

WOUNDS AND PATHOLOGIES

Date: 26 September 2005 (updated 3 July 2006 and 24 Sep 2009) for babase_pending version
Major revision date: 10 September 2020 for Babase version

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 five related tables and four associated views, accompanied by three support tables unique to wounds and pathologies, as well as the bodyparts table, which also supports the darting tables:

1. **Wound/pathology Reports Table** (wp_reports) – Provides basic summary information about all wounds and/or pathologies reported for a baboon on the same date and time. One row corresponds roughly to one wounds and pathologies datasheet.

2. **Wound/pathology Details Table** (wp_details) – Lists specific wounds and pathologies associated with a row in the index table; provides additional specifics. One row corresponds to each unique combination of wound or pathology code and associated specifics (such as impairing locomotion) for a given report.

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

4. **Wound/pathology Heal Updates Table** (wp_healupdates) – Documents the healing status and dates for each reported wound/pathology. One row corresponds to every date of a follow-up comment for each relevant row in reports, details, and/or affected parts (depending on the specificity of the heal update).

5. **Wound/pathology Observers Table** (wp_observers) – Documents the full list of observers listed as initially writing up a wounds and pathologies sheet.

6. **Wound/pathology Reports-Observers View** (wp_reports_observers) – Contains one row for each row from the reports table with an observers column containing a list of the observers who wrote the wounds and pathologies sheet.

7. **Wound/pathology Details-Affected Parts View** (wp_details_affectedparts) – Links specific wounds and pathologies to the body parts affected.

8. **Wound/pathology Heals View** (wp_heals) – Links healing updates to all associated reports, details, and/or affected parts rows.

9. **Wounds/pathologies View** (woundspathologies) – Links associated reports, details, and/or affected parts rows without healing updates.

10. **Wound/pathology Report States Support Table** (wp_reportstates) – Enumerates the possible values (open or closed) in the reportstate column in the reports table.

11. **Wound/pathology Woundpath Codes Support Table** (wp_woundpathcodes) – Defines the wound or pathology indicated by each woundpathcode in the wp_details table.

12. **Wound/pathology Heal Statuses Support Table** (wp_healstatuses) – Defines the values in the healstatus column of the wp_healupdates table.

More detailed information about particular columns can be found in the Babase technical specifications. Here we will focus on how to upload data to and use data in the wounds and pathologies tables and views.

Note that the wid column in reports is assigned by the database manager upon receipt of the initial wounds and pathologies sheet, while the other id columns are all autogenerated by Babase. The wid is used by the database manager to track healing updates as new copies of wounds and pathologies sheets come in with new healing updates and to link pieces of the dataset across the three Excel sheets used for recording and uploading data. For wounds and pathologies that appeared in the babase_pending version of wp_reports (then known as wp_index), the wid remains the same, which is helpful for those with historic wounds and pathologies datasets.

Wounds and pathologies data are uploaded to the following views:

wp_reports_observers
wp_details_affectedparts
wp_healstatuses

This ensures that the various pieces will be properly linked by their various id columns when the data are separated out into their respected tables. Note that, although the wid column is used for uploading to all three views, it only appears in the views and wp_reports, while the other tables link to wp_reports through the wprid column, or in the case of wp_affectparts to wp_details through the wpdid column. Because wp_affectedparts does not connect directly to wp_reports, which means, in the somewhat rare event that you have two distinct wounds or pathologies affecting the same bodypart you cannot attach both to one row in wp_affectedparts but must have two separate rows in wp_affectedparts (e.g., there is a puncture and a slash on the right forearm so there are two rows in wp_details, each with a distinct wpdid and one bodypart cannot be attached to both of those wpdids so two otherwise identical rows must appear wp_affectedparts, one for each related wpdid). The reverse, however, is not true – you can attach more than one bodypart to a given row in wp_details, such as when a large slash cuts across more than bodypart or white fur occurs in more than one location (other than the whole body).

Important

Here we should take a moment to define what constitutes a distinct wound or pathology for purposes of garnering a row in wp_details. Obviously if a wounds and pathologies sheet indicates more than one woundpathcode, there will be more than one row in wp_details. Sometimes, however, more than one row can be generated with only one woundpathcode. This occurs when there is more than one wound or pathology of the same type and some of the supporting details about those separate incidences of the wound or pathology differ. For example, if there are two slashes, one on the right forearm and one on the left thigh, and these two slashes have different values in one or more of the following columns of wp_details: maxdimension, impairslocomotion, or infectionsigns, then two rows are required, one for each slash. By contrast if both slashes would have the same values in all these columns, even if they affect different bodyparts, they should be entered on one row in wp_details. Each row for a given wid in wp_details also gets a value in the cluster column, which is simply a count of the number of details rows for each report. Always begin with 1 and count upwards from there. As you'll see below, this allows the healing updates to be linked to the appropriate wp_details rows (and wp_affectedparts rows where applicable) more easily (i.e., without having to know the wpdid and/or wpaid).

It is also important to note what constitutes a report. Occasionally two wounds and pathologies sheets are created for a baboon on the same date. If they were also recorded at the same time and thus, as far as we can tell, begun at the same time, then the two sheets should be combined into one report. More rarely a second, fresh wound or pathology might be recorded later in the day. In this latter case the sheets should not be combined since they clearly did not originate at the same time. On the other hand observers also occasionally add a new but related wound or pathology to a sheet later. This often happens, for example, with white monkey syndrome. The white fur (woundpathcode 15) tends to be noticeable before a stiff walk (woundpathcode 10) so a stiff walk might be added to the sheet later. At other times observers start a new sheet for the stiff walk. Since these are already long, complicated reports, just go with whatever the Team did.

In the reports tab of the wounds and pathologies Excel file you'll enter the same columns (minus the wprid key assigned by Babase), albeit in a different order. Assign a wid to each new wounds or pathologies report, then enter the sname, date, time, observers, grp, reportstate, and observercomments (which should include the comments section of the wounds and pathologies sheet and any other descriptive details available from demography notes or small notebook pages). When multiple observers are listed, they must be separated by a "/" so they can be separated out into different rows in the wp_observers table and checked against the observers table to be sure the observers exist. Occasionally someone (such as a graduate student) who is not in the observers table will be listed as an observer on a wounds and pathologies sheet. Note this in a comment box in the Excel sheet but do not include that observer for upload. If the report does not yet have a final healing update (any of 3, 4, 5, or 6 – healed, condition permanent, animal missing, or terminal), the report state should be 0 for open. If a report already has a final healing update upon upload the report state should be 1 for closed. Early wounds and pathologies sheets that did not have systematic healing updates or others that are incomplete for some reason (such as a group being dropped while the sheet was active) should also be listed as closed. Open reports are tracked in Excel both to make sure healing updates continue to be sent by the Team and so they can be updated to closed once they are complete.

The ABRP first started tracking wounds and pathologies on standardized sheets in April 1982. There are a small number of wounds or pathologies reported in note form prior to this, which are not included in the wounds and pathologies dataset but should be included in demography notes. Descriptions and healing updates for these early wounds and pathologies might also be found in demography notes rather than on the wounds and pathologies sheets themselves; this is the case

until April 1983. In later years dramatic injuries might also have additional descriptions and healing information in demography notes. In earlier decades copies of wounds and pathologies sheets were only sent Stateside once they were closed but since approximately late 2014 copies have been sent monthly whether a sheet is complete or not. At this time the Team was also instructed to mention wounds and pathologies in demography notes and during monthly data checking the Princeton database manager ensures that sheets have been received for all wounds and pathologies mentioned therein.

For upload to wp_details_affectedparts, the following columns should be filled in: wid, bodyregion (equivalent to the bodypart code listed on the wounds sheet), bodyside, innerouter (more on that in a moment), woundpathcode, cluster (as noted above, you create a cluster number for each unique details row associated with a given wid, starting with 1), maxdimension, impairslocomotion, infectionsigns, detailnotes, and quantity_affecting_part. Bodyregion is used here because the body part # is then combined at upload with the bodyside and innerouter columns to make a comprehensive bodypart code, as defined by the bodyparts table shared with dartings. Body side is L (left), R (right), occasionally C (center), or N (none, not indicated) when no body side is indicated or the bodyregion does not have sidedness. Inner versus outer is not systematically recorded on wounds and pathologies sheets but is an option in the bodyparts table because of dartings so we decided to include it as an option in the wounds and pathologies tables. Most often it will be N (none, not indicated) but when it is clear from the description that a wound was on the inner or outer portion of a body part (such as the inner thigh) then I or O should be indicated. If no maxdimension is provided by the observer, leave this column blank. Impairslocomotion and infectionsigns are not quite binary Y or N columns as U for unknown or unclear is also an available response, which should be used when a yes or no cannot be reasonably determined from the information provided.

For more complex wounds and pathologies records, determining how many distinct details rows are needed requires attention to several details. Remember that you are uploading to the wp_details_affectedparts view, which will then separate the data into two different tables, one for details and one for affectedparts, and that for determining clusters you must focus on the details portion of this data. Obviously when more than one woundpathcode is provided, more than one details row and thus more than one cluster will be needed. (Note, however, that the Team often marks limp along with a wound type when a baboon is limping from a wound and it is long-standing practice not to include limp as a woundpathcode but to instead mark Y for impairslocomotion on the details row for the wound.) Additionally if other columns in the details section differ for a given woundpathcode then additional details rows and thus clusters are needed. For example, if there are two punctures, one on the left ear and one on the right hand, and the one on the right hand is causing the baboon to limp, while the one on the ear obviously does not impair locomotion, then one details row with woundpathcode 2 and impairslocomotion as Y and another with woundpathcode 2 and impairslocomotion N is called for. Likewise (though far more rare) with signsofinfection. And if different maxdimensions are provided for different wounds then those each will require a distinct row in details even if they are otherwise identical. Note that the Team does not always remember to check impairs locomotion and (especially) signs of infection – so long as impairment of locomotion or signs of infection is indicated somewhere on the wounds sheet for a given woundpathcode-affectedpart combination, then Y should be indicated for the appropriate column. The Team nearly never checks no for either of these but provided they haven't checked yes and yes is not indicated by the description, mark them N.

Sometimes the observer chooses a woundpathcode that does not entirely match the description, fails to select a woundpathcode, or later updates to the sheet add further clarity to the extent or

severity of the condition and indicate a different woundpathcode from the one originally chosen. For incoming wounds and pathologies sheets the database manager should ask the Team for clarification on any discrepancies, missing information, or things that are unclear. However, here is a list of conditions where the woundpathcode might not be completely obvious from the wounds and pathologies sheet but that a standardized code is used for. The first five wounds codes are pretty straightforward, though a heavy tick load should be marked as 3 (a scrape or amorphous wound). Other wounds (6) includes things like dislocations, stings, and loose teeth. Since this is a broad category, an appropriate keyword should be added in the detailnotes to indicate what type of wound it is (and this too could lead to two details rows with separate cluster numbers if, for example, a baboon has a loose canine and a dislocation). Often for something like a dislocation, the observer will indicate it might be a dislocation or the database manager will flag it as a possible dislocation – in both of these cases the detailnotes should say "possible dislocation". "Possible" should be used whenever there is some doubt expressed. Limp (7) is also quite obvious, though it should be noted that the current description, "Limp, no wound visible" has not always been the wording. The observer might check it even when a wound is visible but that should be disregarded unless the limp and wound are on different limbs or the wound appears to be very minor. Respiration (8) and digestive (9) problems are also straightforward and, when indicated, the specific type of ailment should be noted in the detailnotes. Malaise, weakness, stiffness in absence of a wound (10) is another one where it is helpful to mark the specific ailment in the detailnotes and it includes things like walking with a hunched back, lying down a lot, experiencing overall weakness or weakness on one side of the body or in specific limbs that results in strange walking patterns different from a normal limp. The stiff walk of white monkey syndrome also belongs here and white monkey syndrome should be indicated in the detailnotes. The stiff walk is most often associated with the hind limbs but sometimes the observer indicates the sacral region instead, which is also acceptable. Thinning fur (11) is used for both comments about fur being thin (sometimes in combination with a change in color) and about bare patches. Note that observers will often mention that a baboon has a rough coat – this is more of an indication of overall health than something wrong with the fur itself so it should just be noted as being present or as improving where appropriate. Nosebleed (12) is used rarely but is self-evident. Discharge, sores, and rashes (13) again should be noted in detailnotes. Other pathology is one that observers often don't check though it applies to a variety of conditions including low body weight, blindness, stunted growth, other signs of delayed development, and any other mysterious conditions that don't cleanly fit anywhere else such as fur turning red instead of the usual white or other pale color. These should be indicated in the detailnotes and associated with the appropriate body part #s (44 for low body weight and stunted growth, 45 for delayed developmental behaviors, 2 for blindness, etc.). White fur also includes yellowing fur and may be assocatied with the whole body or specific bodyparts. Unknown or indiscernible wound type is used when there is clearly some kind of wound but it cannot be seen well due to location, fur, or blood. Scalping is another one that observers might not remember to mark, likely marking amorphous wound instead, but when there is a large wound on the top of the head this should be specified as scalping. We think these are the result of predation attempts by large predatory birds such as martial eagles. Finally broken bone might also be indicated in cases in which it is not checked by the observer, such as when a limb is partially severed (since clearly bones are broken in this process) or when a limp is very severe and in the end the limb has a bend or is permanently short.

Once all the necessary details rows are worked out, the affected bodyparts must be matched up with the correct details rows. Since one wound or pathology can affect more than one bodypart, this may necessitate duplicating the details portion of your Excel row across two or more rows so that each affected part has a row containing the wounds and pathologies details to which it is attached. For example a baboon with white fur on the top of her head, lower back, and tail would

have three rows (bodyregions 0, 24, and 40), each with the same details (woundpathcode = 15, cluster 1, impairslocomotion N, infectionsigns N). Any other details that might be helpful can be included in the detailnotes column. Examples include when more than one dimension is provided for wounds (in which case enter the largest in maxdimension and then include the full dimensions in detailnotes), when the maxdimension is provided in a unit other than cm (in which case convert the number provided to cm and indicate in detailnotes what was provided), or when key words (such as white monkey syndrome, dislocation, or low body weight) might be useful for finding specific types of wounds or pathologies. Note also that a given bodypart might be affected by more than one wound or pathology, in which case that bodypart will also be entered on more than one row in wp_details_affectedparts, one for each cluster it is affected by (e.g., if the left wrist is broken and swollen, there will be two woundpathcodes and thus two clusters/details rows associated with the left wrist and two rows will appear in wp_affectedparts since the same bodypart will be associated with two different details rows and affectedparts rows link up to details rows rather than to reports rows – Babase knows to give one affectedparts row to each distinct combination of cluster and bodypart). Finally the column quantity_affecting_part should be considered. In most cases this will be 1 – for example most pathologies involve single instances of the condition affecting the baboon and more often than not only one wound of a given type is affecting a given bodypart. When more than one instance is recorded and the exact number is provided, use that number, but if vague descriptors like a few, many, or several are used then leave this column blank and give an indication of multiple instances in the detailnotes column.

The field observers have a diagram that lays out the body part #s for various baboon body parts with parts of the head in the single digits, parts of the forelimb in the 10s, parts of the trunk in the 20s, parts of the hindlimbs in the 30s, and parts of the tail in the 40s (but only going up to 43). To these a few other "parts" have been tacked onto the end for the purposes of managing bodyparts in the database and these are assigned by the database manager as needed so that there is always an affected body part. Speaking specifically of the values in the bodyregion column that otherwise match the body part # provided by observers on the wounds and pathologies sheet, these parts are whole body (44), nonspecific/not applicable (45), and not specified (46). Some of these also have the various combinations of left/right/center and inner/outer associated with them in the bodyparts table. Some pathologies, such as white fur or weight loss, might be apparent on the whole body rather than on specific body parts so 44 should be used. In other pathologies, such as digestive issues or breathing problems, an outer body part is not associated with the condition and 45 should be selected as the bodyregion. When the body part # is unknown (and can no longer be acquired via a query to the Team) because no body part # was entered then 46 should be selected as the bodyregion – if bodyside or even inner/outer is known that can employed with 46.

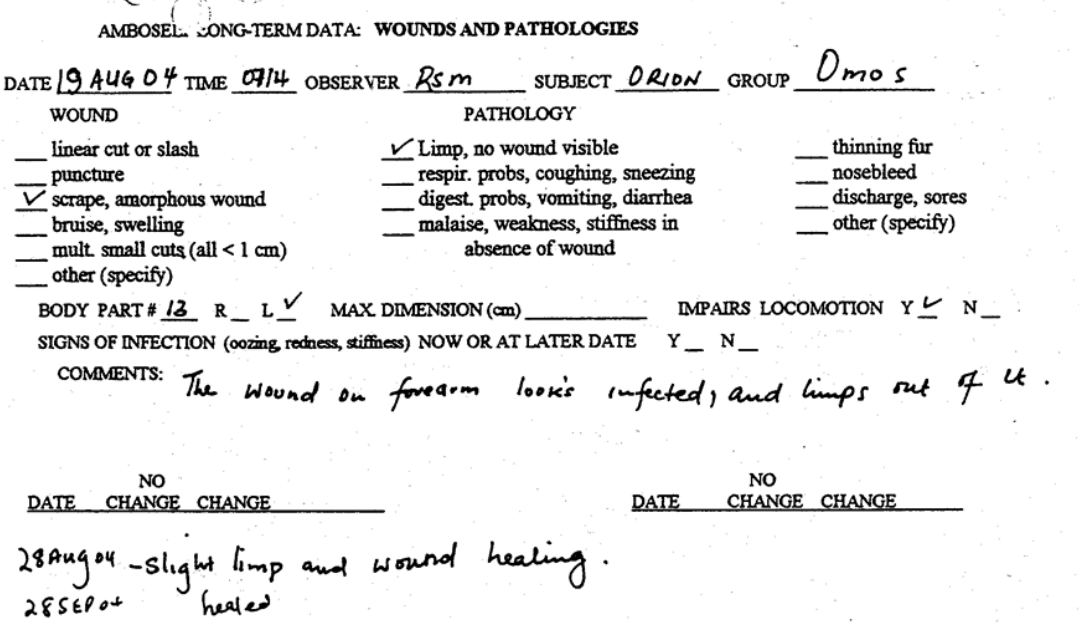
When bodyparts and woundpatchcodes are both known but cannot be matched up with each other there are special rules to follow. For new incoming sheets if this information cannot be ascertained from the sheet, the database manager should ask the Team to clarify. But for older wounds and pathologies reports another remedy was needed since this information is now unrecoverable. These sheets are entered with extra rows using 99 as the woundpathcode to match up to the provided body part #s and using 99 as the unrecoverable bodyregion to match up with the provided woundpathcodes. This results in extra wp_details_affectedparts rows but is necessary in order to link wp_details and wp_affectedparts rows without indicating an association between woundpathcodes and bodyparts that might not be true. Note also that such rows should not be linked to individual healing update rows unless the healnotes indicate some specific information about which wound/pathology or bodypart is or isn't healing (see below).

For uploading to wp_heals, the following columns are available in the Excel sheet: wid, woundpathcode, cluster, bodypart, healdate, healstatus, and healnotes. Of these wid, healdate, healstatus, and healnotes should always be filled in. If the observer wrote any notes about how the healing was going or something about the baboon being missing or dead, that note should be included in healnotes. If the observer simply checked "No Change" after the date in the healing update section of the sheet then write "No change." in the healing updates column. Healstatus is assigned by the database manager based on what the observer wrote in the healing section. A 1 is assigned if there has been no change from the original state or if the baboon started to get better but then got worse to indicate that the wound or pathology is not healing. A 2 indicates the wound or pathology is partially healed. Note that if the status is already partially healed and a new healing update of "No change" is added then the status should again be partially healed. A 3 indicates the wound or pathology is completely healed. A 4, by contrast, indicates that healing is as complete as it is going to get; for example, if a broken limb results in a permanent limp, shorter limb, or a permanent bend in the limb, this would be marked with a 4 instead of a 3. Likewise if a loose tooth falls out, part of an infected tail falls off, etc., these might be considered permanent conditions. (Note: When this dataset was in babase_pending, permanent conditions were considered healed.) A 5 should be used if the baboon is missing from the group and a 6 if the baboon is recorded as dead. Generally guidance on which baboons are missing versus dead is taken from what the observer writes on the sheet. Sometimes both missing and dead may be listed, in which case go with dead. Finally if for some reason a given wound or pathology could not be seen on a given date a 0 may be used. Note that wounds and pathologies sheets in the early '80s until at least mid-1986 often do not have any healing updates because these were not systematically recorded during last days. And those early records that do have healing updates still might not have a "final" healing update of 3, 4, 5, or 6. Note that although 3 and 5 are typically final healing updates, sometime there will be additional healing updates after a 3 or 5 when a wound reopens or a pathology resurges or when a baboon who was missing at the last healing update has returned to the group or turned up in another study group. There can also be repetitions of final healing updates, especially since last days were retooled (such that observers attend each group separately for last days rather than in a group) each observer records a healing update on a different date so there might be two or three (or occasionally more) of the same final healstatus. Finally, since the wounds and pathologies began to be entered monthly, those sheets without a healing update are tracked to make sure the Team continues to update them. There may still be occasional records without final healing updates, such as when a sheet was still active when a baboon's group was dropped or when a sick male baboon is floating among groups.

As for woundpathcode, cluster, and bodypart, ideally each of these can be filled in to match up with each wound or pathology and affected body part with each healing status date. For simpler wounds and pathologies records, this is quite straightforward since there is only one woundpathcode and one bodypart. For more complicated records it the connection of healing updates to woundpathcode and/or bodypart may be less obvious. Generally when there is more than one woundpathcode and/or bodypart and the healing process is relatively brief each combination of woundpathcode and bodypart from the wp_details_affected parts view can still be matched to each healing update such that each combination of wp_details_affected parts rows and healdate will have a row in the Excel sheet. However, occasionally it seems clear from what is written in the healing update that only one wound or pathology (typically the more severe and/or obvious one) is included in the healing update. In that case the database manager may connect the healing update only to the indicated woundpathcode/bodypart combination. Additonally when one or more of the wounds or pathologies recorded in a report drags on for some time, it may happen that the woundpathcode/bodypart combination(s) being addressed varies over time. For example, if one wound or pathology is more severe than another, the less severe one may heal earlier and might not ever be mentioned again since it is already healed. In that case later healing

updates will tend to refer to the more severe wound or pathology and can be assumed to do so even if the healnotes do not specify. Different bodyparts with the same wound or pathology (i.e., when there are multiple wounds of the same type or something like white fur or bare patches on more than one bodypart) sometimes heal at different rates as well so they might all start off as not healed or even partially healed but then one becomes healed while the other is still only partially healed. Occasionally additional bodyparts or even additional woundpathcodes reflecting new symptoms of a condition may be added over time so that a report that started out simple becomes more complicated and begins to generate more wp_heals rows for a given healdate. And when a complicated report is updated for a long time it can sometimes be unclear (especially with simple "No Change" updates) to which condition or specific bodyparts the healing update refers so these types of updates can be applied simply to the wid, leaving the woundpathcode, cluster, and bodypart columns blank. Or, when a healing update clearly applies to a specific wound or pathology but not to any specific parts and several parts are indicated (e.g., such often happens with white fur if the white fur was initially ascribed to several bodyparts rather than the whole body), the woundpathcode and cluster might be filled out while leaving bodyparts blank, at least until specific body parts are mentioned or a final healing status of 3, 4, 5, or 6 is assigned. Finally, as noted above, if 99 had to be used in wp_details_affectedparts because there more than one woundpathcode and more than one bodypart were indicated but with no information on which bodyparts were affected by which wound and pathology types then woundpathcode, cluster, and bodypart should all be left blank unless the observers notes actually indicate to which wound/pathology or bodypart the healing update pertains.

A sample wounds and pathologies sheet.



AMBOSEL. LONG-TERM DATA: **WOUNDS AND PATHOLOGIES**

DATE 19 AUG 04 TIME 0714 OBSERVER RSm SUBJECT ORION GROUP Omos

WOUND	PATHOLOGY	
___ linear cut or slash	✓ Limp, no wound visible	___ thinning fur
___ puncture	___ respir. probs, coughing, sneezing	___ nosebleed
✓ scrape, amorphous wound	___ digest. probs, vomiting, diarrhea	___ discharge, sores
___ bruise, swelling	___ malaise, weakness, stiffness in absence of wound	___ other (specify)
___ mult. small cuts (all < 1 cm)		
___ other (specify)		

BODY PART # 13 R__ L ✓ MAX. DIMENSION (cm) _____ IMPAIRS LOCOMOTION Y ✓ N__

SIGNS OF INFECTION (oozing, redness, stiffness) NOW OR AT LATER DATE Y__ N__

COMMENTS: The wound on forearm looks infected, and limps out of it.

DATE	NO CHANGE	CHANGE
28Aug04		-slight limp and wound healing.
28Sep04		healed

DATE	NO CHANGE	CHANGE

INTERGROUP ENCOUNTERS

Date: 2 November 2009 by Catherine Markham, updated 9 September 2020 by Niki H. Learn

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 individuals 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
    - The Team will record when they spot another group but this does not necessarily infer that this group was near the group they were observing that day (e.g., the group 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)
- During a fission, any group where the subgroups are apart and then come together or are together but seem more like two groups moving side by side than like one integrated group should receive both a subgroup note and an intergroup encounter note such that data users can treat them either as a group in subgroups or as two groups interacting
- 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

Date: Updated 9 Sep 2020

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.

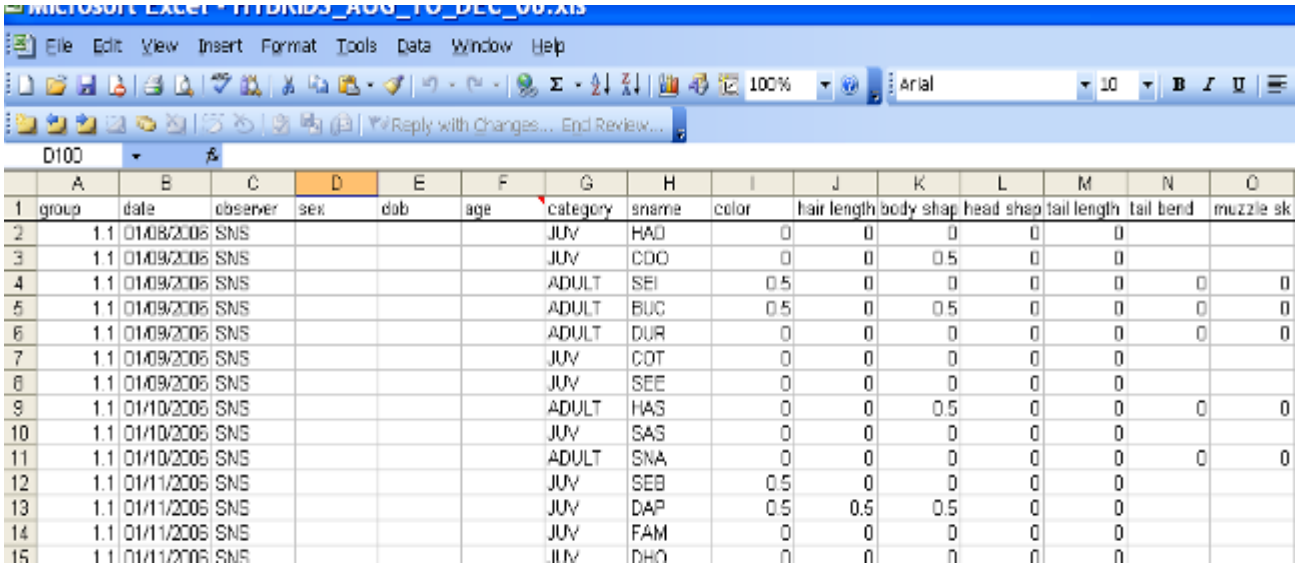
For many years SNS was in charge of setting up the hybridity sheets with the list of which baboons needed to be scored (and for 5 or 7 traits, based on sex and age) each month. Upon her retirement, this duty was passed to ILS but it was soon clear that he was doing a poor job determining whether it should be 5 or 7 traits consistently and nobody had ever transferred the individuals from Laza's group into the Acacia's pages with the Omo's individuals. We also discovered that nobody had been remembering (having no system in place to remember) to rescore juvenile and subadult immigrant males in their birthday months. After many months of trying to sort these issues out retroactively, it became, starting in Dec 2019, the Princeton database manager's job to provide the Team with a list. Typically this is sent after data for the prior month are received, along with corrections to the genetic fecal count list and any errors involving pregnancies relative to what the Team recorded in the monthly report. Females below four years old and males below six years old receive scores only for the first five traits while

females aged four to six years old and males six to eight years old receive scores for all seven traits.

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 as nobody has requested them. The current hybrid_morpho table does not have a column for iterations. Such a 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.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	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	HAO	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 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, updated 9 Sep 2020

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 a few days apart you should receive two neonatal assessment sheets. There are two types, those for the First Observation and those for the Second Observation. The main difference 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 (and wet or not) 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. They also may have useful clues about infants who die due to illness or congenital problems.

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. Sometimes the observer will make qualifying statements about an answer in a field that is to be answered as either Yes or No. For example, in answer to the question "Cling well?" an observer might say "Yes but supported sometimes" or "Not very well, supported most of the time". In the first case the entry for the *cling* column should be Y and the detail that the infant is supported sometimes should be added to the *remarks* if there is not already a similar comment there. In the latter case the answer should be N with a similar note added to *remarks*. *Barespots* is also a Y or N since the answer provided is usually "None". If, however, there were indeed bare spots, the answer should be Y and any description of the size, location, or number of bare spots or a reason for why the bare spots are there should be added to *remarks*. Anything else unusual, such as not being able to evaluate vocalizations because the visit was short should also be included in *remarks*.

Note also that "closed" is an acceptable response in the column *eyes*, though it does not appear as an option on the form alongside clear, cloudy, and crossed. Sometimes when the infant is very new or is sick or weak the eyes will not be seen because they are closed the whole time the observer is watching.

DIGITAL PHOTOGRAPHS

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

Note: Photos are now archived in Photo Organizer, accessible on the wiki. Photo Organizer is managed at Duke; however, below you can find a historical account of how the photos were managed at Princeton in Portfolio, including the naming conventions used in the files passed along to Duke for inclusion in Photo Organizer.

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 may still need to be made, including adding dates for reproductive data collected on incomplete census days, deleting dates when an animal's reproductive data were not collected on that day, etc. 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 a later 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 1998 are not currently available in Babase but older data will be backfilled as opportunity allows.

Getting Started – Creating an Entry Template

Refer to Appendix 3 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 integers 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).

Keep an eye out particularly for small swellings and color changes near the beginning or end of a month as these are commonly missed when the rest of the graph has a uniform appearance.

Typically a full year of sex skin data are entered, proofed, and uploaded sometime in the subsequent year. This allows for all demography data to already be in Babase and for many of the immature females to mature.

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;
```

The Babase changelog

Outside of regular updates, all but minor corrections to notes or tiny tweaks to census that don’t affect any other datasets should be logged in the Babase changelog on the wiki: https://papio.biology.duke.edu/babasewiki/BabaseChangelog

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. Users looking for information about which groups were study groups and when are generally encouraged to use the groups_history view rather than the groups table itself. This section deals with entry and updating of columns in the groups table.

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). For fissions the gid of the parent group should be recorded in the from_grp column in the rows for the offspring groups. For fusions the gid of the daughter group should be recorded in the to_grp column in the rows for the parent groups.

Each group will have various 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 (but “unknown” groups ending in .9 that are used when we are unsure of which fission product was involved must be NULL for permanent date since they are not real groups).

For the oldest groups, this date indicates when we began monitoring them. For fission products it indicates when the group became an reasonably independent and fairly settled group. That is, there may still be floaters but group membership is largely settled and the groups are generally apart and behaving more like separate groups than like subgroups.

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 recorded as belonging to 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.

The study_grp column records the date the group became a study group. For original study groups this will be equal to the permanent date while for fission and fusion products it will be equal to the start date.

Finally, the last_reg_census date column is used 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 usually 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.

Nonstudy groups clearly will not get study_grp or 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) or to prevent data from being mistakenly applied to a parent group that surely no longer exists (as in the case of Kelly's group).

THE WARNING SYSTEM

The warning system consists of a set of queries in the table integrity_queries that are designed to flag errors and potential errors that for one reason or another cannot or should not be kept out of the database upon upload. These queries are automatically run by the warning system on the first day of the third month of each quarter and the results are added to the table integrity_warnings. These should be monitored periodically so that errors might be resolved quickly. Sometimes quirks in the database throw a lot of rows that are not errors at all, such as only having ADF ranks for the original groups, which results in a ton of rows for immature females when the when the MISSINGRANK_FEM query is run. But this query also returns rows for females who are missing from the ranks in a month where they were present, including for some adult females in those original

groups who weren't quite dead yet. When queries such as these are first introduced, it can be time-consuming to work through the initial batch to determine which ones may require action versus which can be "resolved". WARNING: Prior results in integrity_warnings are supposed to simply hang on if they've been resolved or deferred but we have on at least one occasion experienced a situation in which some results were reset with a new value in the first_seen column and with all information in the resolved, deferred_to, category, and notes columns wiped out. It is therefore a good idea, especially for a time-consuming one, to keep a record of all queries run on the integrity_warnings table so that these might be rerun if the table is again reset. Since we are unsure what caused this reset we have resolved to try to track changes to the queries in the changelog in case that is somehow involved.

Some warnings pertain mainly to Princeton data and others mainly to Duke data. The Princeton database manager is responsible for the warning queries of type BIOGRAPH, CDATE, CENSUS, CYCLES, DISPDATE, HYBSCORE, MATDATES, PREGS, RANKS (for females only), SEXSKINS, WEATHER, and WOUNDS, as detailed in the following table.

iqname	type	responsibility
ACTOR_AND_ACTEE	ACTS	Duke
ACT_DIFF_GRP	ACTS	Duke
ADLIB<LBIRTH	ACTS	Duke
CONSORT_OVERLAP	ACTS	Duke
C_STARTS_EARLY	ACTS	Duke
C_STOPS_LATE	ACTS	Duke
INTERXN_NO_CENS	ACTS	Duke, possible need to consult with Princeton
MPI<LBIRTH	ACTS	Duke
BIRTH-LB_TOOFAR	BIOGRAPH	Princeton
BIRTH>MOM_STAT	BIOGRAPH	Princeton
BSTATUS_TOOBIG	BIOGRAPH	Princeton
DCAUSECONF_OF_4	BIOGRAPH	Princeton
EB-BIRTH_TOOFAR	BIOGRAPH	Princeton
EB-LB_TOOFAR	BIOGRAPH	Princeton
ENTRY<LBIRTH	BIOGRAPH	Princeton
ENTRYTYPE_NOT_B	BIOGRAPH	Princeton
RESDNT_B4_ENTRY	BIOGRAPH	Princeton
OLD_CONSORTDATE	CDATE	Princeton
1ST_GRP<>MATGRP	CENSUS	Princeton
ABS_AFTER_CTE	CENSUS	Princeton
ABS_B4_START	CENSUS	Princeton
CENSUS<LBIRTH	CENSUS	Princeton
CENSUS_B4_ENTRY	CENSUS	Princeton
D_CEN_WO_DEMOG	CENSUS	Princeton
NO_PARENT_CEN	CENSUS	Princeton

SNAME_WO_CENSUS	CENSUS	Princeton
CYCLE_WO_POINTS	CYCLES	Princeton
MISSING_SEQUENC	CYCLES	Princeton
MISSING_SERIES	CYCLES	Princeton
NULL_CYCLES_SEQ	CYCLES	Princeton
NULL_CYC_SERIES	CYCLES	Princeton
SEQ_NOT_BEGIN_1	CYCLES	Princeton
SERIES_NOT_BEG1	CYCLES	Princeton
DARTING<LBIRTH	DART	Duke
DISPDATE<LBIRTH	DISPDATE	Princeton
DISPERSE<MATURE	DISPDATE	Princeton
FOC_SAMP<LBIRTH	FOCAL	Duke
MINSIS_PNTIDS	FOCAL	Duke
NEED_FPOINTS	FOCAL	Duke
NEIGHBOR<LBIRTH	FOCAL	Duke
NGH_OWN_INFANT	FOCAL	Duke
SAMPGRP_EXISTS	FOCAL	Duke
HYBSCORE<LBIRTH	HYBSCORE	Princeton
EARLY_MATURITY	MATDATES	Princeton
LATE_MATURITY_F	MATDATES	Princeton
LATE_MATURITY_M	MATDATES	Princeton
MATURE_B4_ENTRY	MATDATES	Princeton
CONCEP<>RESUM-1	PREGS	Princeton
LONG_CURR_PREG	PREGS	Princeton
PARITY_ORDER	PREGS	Princeton
ADM_NO_RNKDATE	RANKS	Duke
MISSINGRANK_ALM	RANKS	Duke
MISSINGRANK_FEM	RANKS	Princeton
RANKED_B4_ENTRY	RANKS	Princeton
RANKS<LBIRTH	RANKS	Princeton
RANK_B4_ENTRY	RANKS	Princeton for ALF/ADF, Duke for ALM, ADM
UNRANKED_OLD_M	RANKS	Duke
SIZE>5_BEFORE_T	SEXSKINS	Princeton
SIZE_0_BTWN_T&D	SEXSKINS	Princeton
LONEANIM<LBIRTH	SWERB	Duke
WEATHERHAWK	WEATHER	Princeton
HEAL>>STATDATE	WOUNDS	Princeton
LATE_HEALDATE	WOUNDS	Princeton
MISSING_AFPARTS	WOUNDS	Princeton
MISSING_DETAILS	WOUNDS	Princeton
REPORT>GRP_CTE	WOUNDS	Princeton
WPREPORT<LBIRTH	WOUNDS	Princeton

There are three basic options for how to deal with a warning row. If the row represents a real error then the error should be corrected. The warning will in turn disappear the next time its warning query is run. If the row represents something that cannot be dealt with right now, such as historic data for a period where demography backfill has not yet been completed, the warning can be deferred by placing a timestamp in the deferred_to column to remind you next time you look at the warnings that this one need not be dealt with at this time. If the warning flags legitimate data that's just not normal for some reason (such as a verifiably late maturity date or those many rows from early rank data where immature females were not ranked) then a timestamp should be placed in the resolved column to mark this row as not being a problem.

In addition to filling in the resolved or deferred_to column, there is the option to add a category and/or notes. Categories currently in use can be found in the warning_remarks support table. Generally if you have many rows that are resolved or deferred for the same reason you will want to apply a category to these rows. For unusual single cases or if you have notes to add, you can leave category blank and fill in the notes column.

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 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\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 Babababase 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 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 infant sname, infant name, birthdate, infant sex, and matgrp for each new birth. The first three letters of the pid 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 bstatus for each birth will be "0" except in the most unusual cases (0 means "known within a few days").

For live individuals the statdate needs to be temporarily entered as the birthdate for entrytype B individuals or as the entrydate for entrytype I or O individuals for the purposes of the upload. It will auto-update in Babase as soon as the new census data are uploaded.

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

Note: Enter the true status, dcause, dcausenatureconfidence, and dcauseagentconfidence onto the update sheets, but be aware that in order to upload new births, you have to temporarily assign a value of "0" for all four columns if that infant's sname shows up on the census sheet anywhere. If you try to upload census information for a status = 1 (dead) individual, the program complains that a dead individual cannot be censused. In the case of an abort, however, the true values can be entered for these columns since the fetus will have no rows in CENSUS.

Finally, assign the infant's full pid 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 11 (unknown) is common for aborts)
Enter the dcausenatureconfidence and dcauseagentconfidence (which will both be 0 if dcause is 11)

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. NOTE: The rare immigrant female is entered using the same rules, except of course that her sex is F.

Enter matgrp as 9.0 ("unknown"), unless observers present some evidence that the male is natal to a known group, in which case that group can be assigned, along with an appropriate matgrpconfidence based on the strength of the evidence (per the confidences support table).

Previously the entrydate for immigrant males was equal to the first date seen in the group according to members but this was changed in Feb 2020 so that like the other left-censored group of baboons (those with entrytype O) the entrydate is equal to the first day present in census.

Enter entrytype as I (for immigrant).

If not assigned, earliestbirth and latestbirth will be autocalculated as birth/2 years before and after birth, respectively. However, individuals with bstatus = 9 must have earliestbirth and latestbirth set to NULL.

For statdate, enter the entrydate. Statdate will auto-update when you load the census data.

For MOST immigrant males

status = 0
dcause = 0
dcausenatureconfidence = 0
dcauseagentconfidence = 0

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

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 pid and parity in the sequence for that mother. To double-check the pid 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 sname and Rdate for every female that Jeanne has marked with an "R". You will not ultimately upload any data from this section because Babase 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.

Pid and parity 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.

Conceive (cpid) 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 sname, and fill in the matured column with the tdate Jeanne has assigned. Mstatus 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 P (puberty) column although it is also OK to leave it blank. This table should all be puberty dates so filling it in is redundant.

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 with the fourth quarter 2015 update (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.

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 sname and the first of the month into the matured 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 matured for Lobo. It happens occasionally that observers have differing opinions in a given month on whether a male is "E" or "Almost E". If this happens wait to record maturity until all observers visiting a

group agree that the male is “E”. It also happens occasionally that a male who has been marked “E” in a prior month goes back down to “Almost E”. The rule is that the male should be “E” for three consecutive months in order to obtain a matured on date. Typically this is not a problem so it is okay to treat a male as matured upon the first recording of “E” but be sure to check whether he remains “E” for the next two months.

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 5yrs, 5mos (the median matage). If the immigrant male is described as a juvenile, he will be added to the scrotal watch list along with the natal males 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 the Duke data manager will decide whether the male should receive an ON or By date, and what date should be assigned. 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.

The database manager annually sends a maturity watch list to the Team with all males and females (in separate sections) who should be added to their respective datasets in each month of the year. Females (unless they have already begun cycling and thus been added early) should be added alphabetically to their group’s sex skin sheets in the month in which they turn 4yo. (In the past these young females were added to the end of the list and not alphabetized until they began cycling.) Males should be added to the scrotal sheet in the month in which they turn 4yo. The Team sometimes misses adding one or more baboons in a given month so always check incoming data against your maturity watch lists so that you can remind the Team to add any 4yos that were missed.

Decision Rules

S. Alberts 4 May 2005
Revised 14 Sept 2005 by LG
Revised 30 Oct 2013 to account for males in dropped study groups
Revised 2 Nov 2015 to correct the median matured by age
Revised 24 Mar 2021 to clarify rule 2c.

Decision rules for assigning "ON" dates and "BY" dates for MATURED (testicular enlargement, onset of subadulthood) and RANKED (attainment of rank among adult males, onset of adulthood) for male baboons in Babase.

- Every male who has been in a study group as an adult will have a ranked date in babase. Every male who has been in a study group as a subadult will have a matured date. Ranked and matured dates will be of two types, "ON" dates and "BY" dates.
- If a date is designated as an "ON" date then we are saying that we know the male attained that marker ON that date (although note that this is not literally true, because we don't track rank changes or testicular changes on a daily basis – males are assigned a ranked date or a matured date on the first day of the month in which we saw them attain rank or testicular enlargement). "ON" dates can be used to estimate the age at which these maturational markers of subadulthood and adulthood are attained. Note that some of the dates from the 1980s and 1990s were not on the first of the month once upon a time but they were reverted to the first of the month ca. 2010.
- If a date is designated as a "BY" date then we are saying that we know the male was adult or subadult "BY" that date but we don't know when he attained it. The point of assigning "BY" dates for ranked and matured is so that we can easily identify which males in any group on any day are juvenile, subadult and adult. The point is NOT to estimate the actual time on which these events occurred, but instead to insure that we have used all available information to know whether a male had reached a given marker by a given time period.
- Note that "BY" dates will NEVER be used to estimate the age at which markers are attained.

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 OR he neither loses nor wins fights with adult or subadult males for several months and then begins to win fights with adult males (and is never seen to lose fights with subadult males), his “BY” date for RANKED and for MATURED is the first of the month he is first present in a study group according to members.

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, as indicated by members.

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 in members 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 (rounded down to the first of the month). 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 the first of the month in which he reaches 7 years 5 months of age (the median age for rank attainment according to Onyango et al. 2013), 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 first of the month in which he enters the non-study group (according to members) 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.

Note: The wording of these rules predates adding residency to members so anywhere it says resident, it just means present per members.

Consortdates

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.

Entrydates

Entrydates used to be a separate babase_pending table but was integrated into biograph in Jul 2016. This table was constructed at Duke but was turned over to the Princeton database manager for future additions as of Nov 2015 since it makes sense to assign entrydates when new animals are added to biograph.

For historical purposes, when entrydates was originally added to Babase, a baboon's entrydate represented one of three dates. For baboons in the original study populations (grps 1, 2, 3, as well as 4 and 5), entrydate is the date they first appear in members (i.e., 14 days prior to the date they were first observed or onset of observation). This also applies to animals entering or born into dropped study groups if they were first seen at least 30 days after the first date on which they were observed (as infants first seen within 29 days of their estimated birthdate will be interpolated present for the entire intervening period and thus must have entrydate equal to birth). For baboons born into existing study groups (or dropped study groups and seen within 29 days of birth, as noted above), entrydate is equal to biograph.birth. For immigrant baboons, entrydate represents the date on which they first appear in a study group (or occasionally a recently dropped, former study group or a frequently observed nonstudy group) **according to members** and thus enter the population of known individuals.

In February 2020 this system was changed slightly to allow for helpful rules when the columns earliestbirth and latestbirth were added to BIOGRAPH (see Appendix 7 for detailed information on assigning these dates). **YOU ARE HERE*** When you can think find a better place to put this so it makes sense or at least make sure its bits are properly referenced in all the places it should be.

Entrytype is equal to
'B' or birth for those individuals who have birth = entrydate.
'I' or immigration for those males with a matgrp > 3.9. These are all cases where entrydate is not equal to birthdate.
'O' for those individuals who were present in a group when close observation began (1 Aug 1971 for Alto's, 2 April 1980 for Hook's respectively, etc.) or who joined a thinly observed group (e.g., infants born during the TZ loss period in the early-to-mid 1990s or infants born to a known mother in a dropped study group that is still sometimes censused where date of first observation is at least 30 days after the estimated date of birth). These are also cases where entrydate is not equal to birthdate.

All baboons with a bstatus of 1 or greater will have an entrytype of I or O, while all baboons with a bstatus of 0 will have an entrytype of B. Baboons with a bstatus of 0.5 may have an entrytype of either B or O depending on whether entrydate is equal to or greater than birth.

Any live individual (any baboon with an sname) must have an entrydate and entrytype. Individuals without an sname (fetal losses, stillbirths, and the rare status 4 individual) should have NULL entrydate and entrytype.

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 in itself adequate evidence to declare a male age 4 and up dead and, lacking other evidence indicative of death, he should be considered censored (which does imply some possibility of death).

If you believe an individual has died, record that individual 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 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:

WAG's estimated death date is 17 Jan 04
– CM 9 Sep 05

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

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 dcausenatureconfidence and dcauseagentconfidence columns 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).

3. But if he was of estimated age (with a bstatus of 1 or greater) he should be assigned a dispconfidence of 0 to signal that his dispersal age should not be used in analyses due to the uncertainty of his actual age.

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 = ddate
R = tdate
P = 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 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. This is rare now that we've switched to quarterly updates but used to happen from time to time when updates covered 6 months of data. You have to be very careful to do a sequential input for females with multiple conceptions during the update 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 during the current update and that cycle has not yet been uploaded. (If the conception actually took place in the prior update period such that the cycle is already in but not the pregnancy, then simply upload the missing preg in Step 3 of the upload process.) If the female conceives and aborts or gives birth during the same update 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.

After you upload all the demography data, developmental milestones, the new cycles and pregnancy data, be sure to update the BIOGRAPH row with the appropriate pid as part of Step 12.

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 separate files to be uploaded at the end of Step 12. Once the first fetus is 'born' in BIOGRAPH (by adding the pid, see above), 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
Major revisions to reflect January 2017 changes: 31 Aug 2020

Criteria for assignment of mortality cause, including the range of evidence (strongest followed by most circumstantial) used for assignment to that category. The dcauses were overhauled in January 2017. This overhaul involved some reorganizing of the dcauses to group them into subcategories known as natures and, in some cases, to break out different agents from a nature. Previously the dcauses were 1 = predation, 2 = conspecific, 3 = wounds and injuries from accidents, 4 = pathology or obvious congenital problems, 5 = loss of mother, 6 = human action (which did include domestic dogs), 7 = unknown, 8 = under review, and 99 = censored. There are now more dcauses with dcause agents nested within dcause natures (primarily violent, pathology, and interruption of maternal care). At this same time the dcauseconfidence column was split into two confidence columns, one for nature and one for agent.

NATURE = VIOLENT

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 now be assigned to dcause 5.

4. *Human or domestic dog*: In earlier years this source of mortality was observed primarily for the groups living around the tourist lodges but in more recent years, beginning notably during the 2009 drought when human-baboon conflicts over goats occurred, it has resurged due to the larger baboon population and frequent influx of new herders who don't know not to mess with the baboons. Evidence ranges from finding an individual with machete wounds to injuries obtained from having been hit by a vehicle to multiple disappearances in an area when many people and dogs were about, often combined with confirmatory reports or rumors in the lodge worker or the Maasai community, such as a sudden disappearance combined with reports of an animal being so injured.

5. *Unknown violent*: Any case where we're sure the baboon was injured and we think that injury likely led to the baboon's death but we have no evidence suggesting one particular agent over another should be assigned this dcause.

NATURE: PATHOLOGY

6. *Infectious pathology*: 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 prevalent illnesses, our records include several instances of apparent epidemics, some involving paralysis and likely due to coxsackie virus.

7. *Noninfectious pathology*: Evidence ranges from (strongest) obvious congenital problems or severe weakness in a single individual with no apparent injuries through (more circumstantial) a disability including prolonged limping that persists after an injury has otherwise healed or worsens in old age with arthritis. 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. 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 if kidnapping is not evident.

8. *Unknown pathology*: As it is often difficult to determine the cause of illness from the available information, this dcause is also available for cases where we're certain a pathology was involved but have no confidence as to whether the agent was infectious or not.

NATURE: INTERRUPTION OF MATERNAL CARE

9. *Loss of mother*: This category includes instances where an infant disappeared simultaneously with or very shortly after its mother and the infant is believed to have died due to lack of maternal care. Loss of mother is used for all cases where an infant died in utero upon the death of the mother.

10. *Other interruption of maternal care*: This dcause is used to include other instances in which maternal care is interrupted for a prolonged period without involvement of a kidnapper, such as getting lost from the group despite having a live mother (e.g., Moon) or where the mother was severely injured and unable to care for the infant (e.g., Lohud).

UNKNOWN

11. *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).

UNDER REVIEW

12. *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.

NOT APPLICABLE

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 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, which is rare but has happened.)*

In 2011 we 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
- 1 – Possible, with limited evidence
- 2 – Likely, with circumstantial evidence
- 3 – Very likely, with corroborating evidence
- 4 – Confirmed (used only for dispersals)
- 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).

See the section ***Dispersal Dates for Natal Males*** above for more information on assigning dispconfidence values.

Initially the confidence level for a dcause was encapsulated in a column called dcauseconfidence but when dcauses were overhauled in January 2017, the confidence was split into two pieces, which can be found in the dcausenatureconfidence and dcauseagentconfidence columns of BIOGRAPH. 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.

The dcausenatureconfidence column describes our confidence that we have chosen the correct nature for the dcause. The available natures are violent, pathology, interruption of maternal care, unknown, and under review. These are the subheadings in the list of dcauses above with each dcause representing a particular agent. Our confidence that we have chosen the correct agent is then described in the dcauseagentconfidence column. Since agents are nested within natures, the dcausenatureconfidence can be higher than the dcauseagentconfidence whenever we are more sure of the nature than of the agent.

A confidence of 0 is assigned when a baboon is still alive, when a baboon is censored, or when the dcause is 11 (unknown) or 12 (under review). Thus if the status is 0, 2, or 3 or if the status is 1 and the dcause 11 or 12, then dcausenatureconfidence and dcauseagentconfidence should always be 0. Also if the nature of a dcause is known but not the agent such that dcause 5 (violent in nature but with unknown agent) or 8 (pathological in nature but with unknown agent) is employed then the

dcauseagentconfidence should be 0 regardless of the confidence level assigned to dcausenatureconfidence.

A confidence 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 dcausenatureconfidence and dcauseagentconfidence of 1 should be differentiated from other cases when examining dcauses. Note, however, that one column might be a 1 while the other is higher, if for instance the nature is likely or definitely of a particular category (violent, pathology, etc.) but the precise agent is less certain.

A confidence 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 dcauseagentconfidence of 2 may be assigned (along with a dcausenatureconfidence of 3 in the described case).

A confidence of 3 is assigned when there are multiple lines of corroborating evidence such that the assigned cause of death is rather likely or there is little-to-no doubt.

A confidence of 4 is not assigned because it was difficult to discern what other than direct eyewitness accounts constituted a confirmed dcause. Therefore, any case that was previously considered to have a confirmed dcause was reassigned to dcausenatureconfidence and/or dcauseagentconfidence 3.

APPENDIX 3: CREATING A SEX SKIN DATA ENTRY TEMPLATE

C. Markham
Revised by Tfenn: 24 April 2009 and later by N. Learn

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 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

```
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 = 'C' order by extract(month from date),
census.sname, extract(day from date);
```

The = 'C' portion is very helpful, as it reduces messiness in the dates grabbed from members by removing manual census points on the estimated date of death, making it easier to tell when a female was seen during a group fission by not counting 'A' rows, and generally not counting rows when a female was absent.

Alternatively, you may wish to use `census.status <> 'A'` if you are working in a period with members-style census data. This will then only filter out the dates when the female is marked absent.

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

sname	date	size
CHE	5-Jan-99	0
CHE	7-Jan-99	0
CHE	9-Jan-99	0

PCS KINS

sname	date	color
CHE	5-Jan-99	B
CHE	7-Jan-99	B
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 your 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. These sex skins cannot be entered. Check with Jeanne regarding juvenile females who have mini-cycles while still in a pre-pubescent phase.

APPENDIX 4: 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.

APPENDIX 5: CYCGAPS by Niki H. Learn

Cycgaps is a strange beast of a table. As the name indicates, it is the place to indicate when a gap in cycling data has occurred but the gaps it details are really indicative of reproductive gaps, not merely cycling gaps, since we may have gaps in the reproductive information for females who are not in fact cycling during part or all of an observation gap. As a result, things got complicated.

Although the table originally seems to have been focused only on gaps in cycling (as indicated by both its name and the historical end and start of observation rows…or lack thereof for females who did not cycle at all during the observation gap) yet it impacts what the database tells us about reproductive states other than cycling.

Both cycstats and repstats (the daily records of a female's cycling and reproductive states respectively) have gaps when there is a cycgap in place. That is, if there is a gap in cycling, there will be no rows in either cycstats or

repstats for the dates between the end of observation row and the next start of observation row (or after the end of observation row if there is no matching start of observation row), not inclusive. (Except of course where there is a cycgaps point row, indicating a single observation point within a gap, as when a group is observed during an otherwise gappy period or after the group is dropped.)

Additionally, one of the cycgaps columns is the female's state on the date of the row and the options for this state are a mishmash of both cycling and noncycling reproductive options:

M	menses	follicular -- Mdate (inclusive) to Tdate (exclusive)
S	swelling	follicular -- Tdate (inclusive) to 5 days prior to Ddate (exclusive)
O	ovulating	5 days prior to Ddate (inclusive) to Ddate (exclusive)
D	deturgesence	luteal -- Ddate (inclusive) to Mdate (exclusive)
P	pregnant	Ddate (inclusive) to birth (exclusive)
L	lactating	birth (inclusive) to Tdate (exclusive)

Unfortunately many of the older cycgaps in Babase were focused on the cycling aspect such that some gaps in cycling are shorter than the actual gaps in observation (because a female was cycling for only part of the observation gap) or are missing altogether (because the female was known to be pregnant and or in post-partum amenorrhea, PPA, throughout the observation gap). This is problematic because there are a lot of rules for cycling data (rightly so to help prevent errors from being introduced to the cycling data) that make it difficult to edit cycgaps, particularly because transactions were not available in Postgres when these tables were designed and the trick that's supposed to allow the addition of new cycgaps within a period of observation does not presently work.

As a result there are a lot of incorrect data in repstats in cases where there was a gap in observation but that gap is not reflected in the cycgaps table. There are also a number of incorrect (and quite nonsensical) reproductive states recorded in start of observation rows within cycgaps because the rules governing this column work well if cycgaps rows and additional data after observation starts up again are added sequentially but not so well if changes are attempted at a later date. The current rules also do not allow for a female's sex skins to be entered for a cycle that was picked up at any time after the mdate when observation resumed (which is a problem for some Omo's members now in Acacia's group).

Jake and Niki worked on some potential fixes for these problems in fall 2015 but most of these have not yet been implemented because of objections Karl made, though at the 2016 joint lab meeting the leaders and database managers decided that these fixes were far preferable to rewriting the code from scratch. For now the database manager should be aware of these issues until such time as they are fixed and corrections can be made.

Adding new cycgap rows in real time:

When a group is dropped all females 4yrs or older should be given an end of observation row in cycgaps on the last regular census day. The same should be done if a group or individual is temporarily not observable for a lengthy period of time. Females under 4yo do not need them initially, though any who continue to be observed during other groups censuses through the age of 4yrs should also be assigned end of observation rows on the last regular census day. The state column is supposed to be blank for end of observation rows (though some from the TZ loss period in the '90s are not blank).

When females who are currently in an observation gap are observed and their reproductive state recorded, they should receive a cycgaps points row on that date with the state column indicating their cycling or other reproductive state. Due to the lack of surrounding contextual data, NHL developed the following rules of thumb for assigning points rows:

- Any reproductive information on a female is included in the accompanying demography note.
- The Team's call on whether a small swelling is turgescent or deturgescent when they have no information about the trajectory of the cycle seems somewhat unreliable. Since the cycgaps state must be assigned as either a T or D (or O), swellings below size 3 are not assigned cycgap points rows, though the information is still recorded in the demography note.
- It is impossible to tell for sure whether a female with a turgescent swelling would qualify for a rating of ovulating without additional information so I assign an O if the female has a 7T or larger swelling and/or is noted to be consorting and otherwise assign a T.
- If a female appears to be pregnant (she is recorded as P/B) then she should be assigned a pregnancy point row in cycgaps. However, if data are slim, it can be difficult to differentiate between pregnant females and those who have recently miscarried or have an infant that is not noted (or perhaps even who simply has a lot of pink islands). There may also be very little to go on when it comes to estimating conception and birth dates. Therefore a pregnancy is not added to the pregs table until there is some additional context indicating that an infant was born, or perhaps that a miscarriage or early infant loss occurred followed by another pregnancy. It is important to consider the timing between pregnancies when attempting to estimate these dates when data are sparse.
- Since it is often difficult to tell how old a "brown infant" might be and a female generally resumes cycling while her prior infant could still be called a brown infant, only the presence of a black infant with the female indicates addition of an L (PPA) cycgaps point row. It is quite rare for a wild-feeding female in our population to resume before their prior infant stops being called a black infant.

Of course if regular observation of females begins or resumes after a period of no observation then any females over the age of 4yrs (or less than 4 but already matured) must be assigned a start of observation row in cycgaps. The state should match whatever cycling or other reproductive state the female was in when observation resumed.

Babase will not allow a cycpoints row during a cycling gap unless either a) there is a matching cycgaps point row or b) the source column in cycpoints is E (for estimated). When a group is first dropped or sometimes when observation is otherwise decreased it might be reasonable to assign conception, tdate, or ddate rows in cycpoints with a source of D (data-based) if there is enough data available to do so. These then must have matching point rows in cycgaps. Otherwise, any cycling dates assigned to a female whose estimates are not accurate to "within a few days" should have a cycpoints source of E. WARNING: The source column is not included when uploading cycles for females under regular observation and these rows are given a source of D by default so if you are uploading estimated zdates for females within a cycling gap be sure to include the source column (and any relevant early or late dates) and assign a source of E. In periods of low observation (such as during the TZ loss period) there may be tdates and nonconceptive ddates with source E but usually for dropped groups only conception dates will be estimated.

Estimated cycpoints rows do not need and should not have matching cycgaps point rows. This is because any date that has a row in the cycgaps table (including cycgaps point rows) will also have a row in repstats and (if cycling) in cycstats and we do not want rows there for estimated dates but rather reserve that distinction for information that was actually observed on the occasions the females were seen. Additionally, if for some reason an estimated date needs to be changed (especially a conception date, which is also linked to the pregs table) – say because new information is available or a mistake of some kind has been discovered – it is easier to make changes to these rows if they do not have matching cycgaps rows to move with them.

Estimated conception dates (and birth dates) should, however, have corresponding demography notes, which ensure that the mother (or mother and infant) is marked present in the appropriate group on the estimated event date and can provide an indication of the precision of the estimated date.

NOTE: Normally when uploading cycpoint rows we do not even include the source column in the upload file, consequently the rows are automatically assigned a source of D. If you have rows that ought to have a source of E be sure to include the source column in your upload file. WARNING: Babase has a rule stating that the source column cannot be changed. In 2015 we had to temporarily turn rules off in order to make corrections to older rows and to newer rows once NHL realized that they ought to have been uploaded as E rows. Since you have this guide you can avoid that!

You should also assign early and late dates to estimated tdates or ddates. These additional dates indicate the ends of the range of reasonably possible dates that might otherwise have been assigned as the tdate or ddate. Like the source column, these are additional columns you can add to your upload file when you need to assign them. Note that in cases where uncertainty for a conception date is high you may need to truncate the late date a bit because you cannot have a late date after the birth of the involved infant.

Please see the technical specifications for Babase if you need additional tips on the rules of cycgaps and how to edit cycgaps rows. If you must edit cycling data during a cycgap, brace yourself as you may accidentally rip out large quantities of your own hair.

APPENDIX 6: CYCLE SCORING by Niki H. Learn in consultation with Jeanne Altmann

This appendix is a supplement to Step 1 of the demography update (Manual data review and consolidation, section 1a, Jeanne's Review), describing in detail Jeanne's technique of assigning tdates and ddates. Please see the referenced section above for an overview of what these dates represent and remember to mark out any nonexistent dates (the 31st of several months and the last two or three days of Feb) with the pink pencil to be sure these dates are not accidentally considered as possible tdates or ddates.

Jeanne assigns tdates and ddates in green colored pencil by inserting a downward-facing arrow in the column matching the desired date with the appropriate letter (T, P, or R for tdates or D or Z for ddates) above the arrow.

These cycling event dates are usually not plainly obvious since data are not collected each day; thus they generally must be estimated based on the expected trajectory of the cycle's curve. Note that the deturgescent slope is steeper than the turgescent slope such that a female's swelling increases is size much more slowly than it decreases. The slope can also be variable between and within females as well as generally over time. For example, in the earlier years of the project females tended to deturgesce faster than they have done in later years (be sure to keep this in mind if you ever must backfill any cycling data). As another example, the turgescent slope is sometimes more gradual and drawn out, often during puberty or resumption cycles. Puberty and resumption cycles also may result in smaller maximum swelling sizes. Maximum swelling sizes also vary among females and tend to go down across females during times of severe drought.

It is difficult then to provide hard and fast rules for assigning cycling event dates beyond the exhortation to look at the available data closely, including any

accompanying reproductive notes, and use your head. Beyond that here are some examples and tips for dealing with different types of situations, moving from fairly clear and easy calls to more difficult cases for each event type.

Mdates

Mdates (for menstruation) can be assigned by the database manager and otherwise will be automatically generated by Babase. Mdates are marked on the sex skin sheet by using a red pencil to circle the first M indicated on the female's sex skin sheet after her last ddate. The corresponding date will be her mdate. Note, however, that mdates must be on or before the tdate so the rare instance in which the only M marked is after the tdate, that M must be disregarded.

If no Ms are present, do not assign an mdate, Babase will automatically assign one a set number of days after the ddate (unless this is a puberty or resume cycle in which case there cannot be an mdate associated with the cycle).

Occasionally a female will be seen with blood on her perineum around the time she would normally menstruate despite the fact that she is newly pregnant. These Ms are not entered (and Babase will reject them if upload of such Ms is attempted).

Tdates

When the first sign of a turgescence is a size 0-1 (0.5) swelling, assign a tdate to that date.

When the first sign of turgescence is a 1T swelling, exact placement of the tdate may vary based on the slope of the ensuing turgescence and the number of days between the 1T swelling and the last record of the female having a flat and black sex skin. If, for example, the female was flat on the first of the month but had a 1T swelling on the second then the tdate would be placed on the second. However, if the female were flat on the first and had a 1T swelling on the fourth then the tdate might best be placed prior to the fourth (probably on the third) unless perhaps the swelling increased in size very quickly thereafter. And if the 1T swelling is on the third then it can reasonably be placed on either the second or the third; take your best guess given the slope of the rest of the swelling. If in doubt use the midpoint or flip a coin.

At other times the first nonzero sex skin measurement may be larger than 1T (usually 2T but generally not larger than that unless data are sparse or there is an unusually large number of consecutive nonobservation days in otherwise near-daily data, as occurs occasionally when staff are in short supply or when weather, vehicle problems, or TZ difficulties prevent timely observation). In these cases the tdate is extrapolated backward from that first swelling based on the slope and shape of the rest of the swelling.

Once in a while a female appears to begin a cycle but it never really gets going or it dips down again before progressing to a large swelling.

Very small cycles that never get over 1T are ignored unless they last for a fairly long time (at least 10 days?). This situation is most likely to occur in pre-pubescent females, though it also sometimes occurs when a female has been in PPA. If the cycle does not reach at least 2T and is not sustained for a reasonable length of time then it is not entered into the cycling tables and is not counted as either the date of puberty or the date of cycle resumption. A female's maturedate must match the tdate of her first cycle and her resume date is automatically assigned as the first tdate after the birthdate of her prior infant. If, however, there is a funny little swelling that drags on for a bit but eventually gets larger, score that as you would normally, including the whole of the very small swelling.

Ddates

Rarely a clear case will present, in which there is a large turgescent swelling on one day and the following day is also an observation day but now the swelling is deturgescent. Here the latter date is marked as the ddate.

Occasionally the Team will also indicate in the reproductive notes that a female seemed to have a tight swelling when they arrived but that later it was showing wrinkles. They may also mark this on the sex skin sheet by filling in the open circle they original drew. This is clearly the ddate. Depending on what the rest of the sex skin graph looks like, an indication in demography notes that males stopped paying attention to the female mid-observation or that males are ignoring a female who appears to be turgescent may also signify that the date in question should be the ddate, unless the data in the ensuing days suggest otherwise.

More often you will find one or more noncensus days between the last turgescent day and the first deturgescent day. Here the key is to look at the downward slope and extrapolate it back up to the maximum swelling size to estimate the ddate.

Unusual Cycles

Occasionally females under some kind of stress (or, most infamously, a female called Kathryn who only infrequently experienced normal-looking cycles) will have bizarre-looking cycles.

We have a small number of cases where a female's swelling suddenly dips down in the middle before going back up again and completing a normal trajectory from then onward. These dips are sometimes traceable to an injury, either an injury to the sex skin specifically or a major, system-shocking injury to the female, while in

other cases the cause is unknown. Whatever the source of these dips, they do not affect assignment of tdates and ddates.

Kathryn's cycles on the other hand often went up or down very gradually and the swelling sometimes did not go all the way down to zero between apparent cycles. Conic occasionally experiences this problem also. We do have a small number of isolated cases of this occurring in other females. If there is something resembling a cycle to work with then do your best to pick out where the ends ought to be even if the swelling never quite goes down to zero in between apparent swells in sex skin size.

When data are not near-daily

If cycling data are sparse or gappy, with some luck you may still be able to estimate the tdate and/or ddate for some cycles reasonably well, but at other times you will either wish to have a cycling gap and only enter those cycling event dates that are very clear and any estimated conception dates (if low-frequency observation persists over a long time) or take your best guess at assigning estimated tdates and ddates where needed (if low-frequency observation lasts only for a short time or if the problem is an isolated, smallish gap).

Reminder: Be sure to assign a source of E (estimated) in cycpoints to any tdates or ddates that cannot be known within a few days in either direction due to sparse or gappy data. This assignment should be made during upload as Babase does not allow alteration of the cycpoints source column. Do not estimate mdates, as these dates should always be either taken directly from the data or automatically assigned by Babase.

NOTE: In older periods when data were sparse or gappy sometimes only the ddates were estimated despite the females not being assigned a period of no observation in cycgaps. This has resulted in errors within the day-to-day reproductive tables (cycstats, repstats, etc.). At some point this should be fixed either by estimating the tdates or inserting new cycling gaps (once the Babase code is fixed to allow new gaps within periods of observation).

Assigning conception dates

When it took weeks for data to arrive from the field via mail and updates were done every six months instead of every three months, it was fairly rare to not know whether a female was pregnant at the end of an update period by the time the update was actually completed. Now, however, with scanned field data arriving within days and the quarterly demography update typically being completed before the next month's data arrives, more females may be up in the air at the time of cycle scoring.

A preliminary assignment of pregnant or not is often made for females whose ddate occurred in the last or second-to-last month of the update period without the female either beginning a new cycle or starting to show pink on her PCS. Several things are considered when deciding:

- If the female is nulliparous and only recently matured, she likely is not pregnant, especially if males have not yet been consorting with her. Usually it takes at least several cycles before adult males begin paying serious attention to nulliparous females, though occasionally a female will conceive within the first few cycles. In recent years, females often cycle for more than a year after maturity prior to conceiving for the first time.
- A female who has just resumed cycling is fairly likely to conceive on her first cycle if she just miscarried or lost a young infant but less so if she has thus far successfully raised her last infant and, especially if she has resumed on the early side or is in relatively poor health. In these latter cases she may pause for one to three months before embarking on her second post-resume cycle. We call this a post-resume pause or shutdown (see more on shutdowns below).
- A female is more likely to conceive with each cycle since resuming so a female that has had several cycles may be more likely to have conceived. But beware since this is complicated. Some females tend to conceive on fewer cycles than others. Drought may increase the number of cycles a female goes through before conceiving, whereas particularly good rains may result in earlier conception. Older females may cycle much longer before conceiving and occasionally will stop conceiving altogether though they continue to cycle regularly.
- Young females experiencing their first pregnancy tend to start showing pink later (or perhaps just less noticeably) than multiparous females.
- Consider the number of days the female has remained flat since her last ddate relative to the number of days she is typically flat between cycles. This too can vary over time and among females but patterns are often detectable.

Females who have been flat for a couple weeks or more leading up to the end of the update period may be tentatively marked as pregnant (Z) if the above indicators suggest that she is likely to be pregnant. If it's a bit less likely but she might be pregnant, then she may be marked with a "Z?" or, especially if the last ddate is close to the end of the update period, the letter may be left off altogether (leaving only a question mark) until more information is available. If the indicators suggest that she is not pregnant then she may be marked with a "D?" or if she is unlikely to be pregnant, simply marked with a D.

As long as field communications with the Team are good at the time, it can be beneficial to make a list of the possible pregnancies and ask the Team in the second half of the first month of the next update period whether the females are now showing any pink or are again cycling. If the Team cannot be reached or can but are not yet able to provide definitive answers for all females of

questionable state, then assign pregnancies to those females believed to probably be Z and not to those that are less likely to be pregnant. But be sure to keep an eye on them and confirm both the pregnancies and nonpregnancies when further data is available. It is easy to miss a pregnancy at the next update if those with question marks are not carefully checked.

For females in dropped groups or those with cycling gaps during periods of low observation, we typically only enter estimated conception dates and do not attempt to estimate other tdates or ddates. If the conception date cannot be determined within a few days on either side then the cycpoints source column should be set to E (estimated) at upload and edates and ldates should be assigned to indicate the range of possible conception dates.

If a conceptive cycle was not observed but the infant's birthdate is estimated with high precision then the conception date should be estimated by subtracting the median gestation time (178 days) from the infant's birthdate. These conception dates should still have a source of E, as well as edates and ldates, in cycpoints since gestation can reasonably last up to three weeks less or more than the median. Note also that in the earlier years of the ABRP 177 days was used as the gestation period.

In other cases estimation of a conceptive date may be more difficult but using any known cycling data, any sightings of the female with pink showing on her PCS, and the first sighting of the infant, one can often narrow down the conception (and birth) date to within a few weeks to a couple months. This most often occurs with females in recently dropped groups or females in groups that have been seen fairly regularly over the recent past. Putting all the pieces of evidence together can give a reasonably good picture of when conception and birth occurred. For example, if a female was seen with a large swelling and being consorted in Mar, was flat and black in Apr, was P/B in both May and Aug, and had a black infant with her in early Oct and again in Dec we can deduce that she conceived on the Mar cycle and gave birth in Sep. We would then choose a conception date on or shortly after the Mar sighting.

When data are very sparse, particularly if little is known of the timeframe in which pregnancy and birth occurred (especially as the age of the infant at first sighting will usually be only approximate, i.e., the female had a black infant – a label applied from birth to roughly 4-6 months of age – or a brown infant – a label applied from roughly 4-6 months of age until as late as a year and a half or more), then the estimate may need to be determined using the midpoint between the last date she was known not to be pregnant and the first date at which point she must have been pregnant. This estimate should be tweaked by accounting for variation in gestation time and any available clues.

Shutdown vs. fetal loss

Sometimes a female will cycle, then remain flat and black for an extended period (typically 1-3 months but sometimes longer) before cycling again without ever having shown any pink on her PCS. One must then decide whether the female has embarked on a temporary shutdown or conceived but lost the fetus before she began showing pink.

There are some clues we can use to make this determination. Unfortunately, the presence of blood on the perineum around the time menstruation would normally be expected is not one of them since this sometimes occurs despite females having conceived and of course might also be observed when a shutdown is occurring. However, bleeding near the end of a prolonged period in which a female is flat and black is indicative of a fetal loss and a pregnancy will be assigned.

There are three types of situations in which a shutdown may occur. If we suspect that a shutdown has occurred rather than a brief pregnancy

Females who have recently matured may pause between cycles, particularly between the first and second cycles. In these cases it is clear that the females are not pregnant, especially as males generally ignore females for their first several cycles. An extreme case of this was observed when Janet was severely wounded shortly after maturing and delayed her second cycle until 7.5 months after her first. We can safely assume that these females are not pregnant.

Another type of shutdown is a post-resume shutdown in which the female has a resume cycle, sometimes smallish, other times normal looking, but then takes a little rest before the second cycle. This may be more likely to happen when the female resumes earlier than she typically would, when conditions worsen (such as a delay in the onset of the “wet season” in which case her health may suffer or she may simply need to allocate more resources toward supporting her prior infant), or when the female’s condition is temporarily impaired by a wound or pathology. This type of situation is easy to spot but we cannot be sure that the female has shutdown, as pregnancy and early fetal loss might also be a feasible explanation.

In other cases a female with more than one cycle under her belt since her last resume might shut down temporarily in response to some substantial worsening of her condition such as that due to a severe drought or a serious wound or pathology. These occur more rarely.

We cannot predict with full accuracy based only on the presence or absence of the above circumstances whether or not the female had a shutdown or a pregnancy with an early fetal loss but we have another tool in our arsenal – our fecal hormone data. As it will likely be years after a suspected shutdown or fetal loss before hormone results are available (if we in fact have any samples for the female from the appropriate time period), we must take our best guess as to

whether the case is a shutdown or a pregnancy with early fetal loss and then confirm that guess or make corrections when hormone data become available.

Thus we put these cases on a list for Laurence to look at once hormone results are available. In some cases we will not have any fecal samples during the span of the shutdown or pregnancy and in others we may have samples but they're too close to the potential conception or fetal loss dates to tell us anything one way or the other. But in other cases we have been able to confirm or discount pregnancy and even occasionally refine the date of fetal loss based on the mother's reproductive hormone data. NOTE: When adding or deleting pregnancies after successive pregnancies have been added, remember that pids are never changed but Babase will automatically update the parity for all involved infants in order to keep track of birth order.

Note also that the time from conception to the first sign of pink showing on a female's PCS is variable. The median time to start showing pink following conception is about two months but it is highly variable within and between females such that during a given pregnancy a female might be marked P/B before the end of the first month or after the end of the third. Some females embarking on their first pregnancy, for example, will take longer than average to begin showing pink while others won't. The oldest females tend to start showing pink earlier, though observers need to watch carefully since females who have had many infants often have permanent pink islands on their PCSes that do not turn back to black after parturition. Females also may take longer to show pink in times of duress, such as during a severe drought.

APPENDIX 7: ASSIGNING INFANT BIRTHDATES by Niki H. Learn in consultation with Jeanne Altmann

Assigning infant birth estimates for live births or stillbirths (where the dead newborn is observered) requires examination of the neonatal sheets, any relevant notes, and the time elapsed between when the mother was last seen without the infant and first seen with the infant. To make the latter obvious on the sex skin sheets, Jeanne draws a bracket with her green pencil from the day after the mother was last seen without her infant through the first day the mother was seen with the infant. The dates encompassed within the bracket, including its ends, are the possible birthdates. The pink lines drawn through the 31st of months with 30 days and the last two to three days of February again come in handy to make sure those dates are not considered as possible birth dates.

The first neonatal sheet may offer clues that narrow down or pinpoint the date of birth. For example, a neonatal sheet which states that the infant had a pink or white flexible umbilical cord, that the infant was bloody or had cloudy or crossed eyes, and/or that the mother's mouth, hands, and/or feet are still bloody and wet highly suggest that the infant was born on the first day seen. This does not

happen very often, and even more rarely is a birth observed by the Team, but in these cases we have an obvious choice for the date of birth.

Usually, however, the clues are not as clear and birth date will need to be estimated. Any umbilicus, including a short dark dried one, may still point toward a birthdate on the date first seen, especially if the infant was sighted in the afternoon and the mother was without the infant the prior day and was still somewhat bloody. Otherwise this is likely to indicate a birth on the prior day. After the day of birth the mother is more likely to have blood visible only on the perineum, if anywhere.

Whether or not the infant needs help clinging and is clutched toward the nipple by the mother during the first and second neonatal assessments can also offer some hint. If, for example, the infant is not clinging well at the first assessment but seems fine at the second assessment, especially if the first assessment was in the morning when the infant should not yet be tiring, suggests that the infant was born recently (whereas one that continues to have trouble clinging likely was premature or has a health issue). When the mother is still aiding the infant and clutching the infant to the nipple during the second assessment and indications are that the infant is healthy, this suggests that the infant was born toward the later end of the range of possible birthdates. In contrast an infant who was born during a several-day span between observations and is not (or only rarely) being clutched to the nipple or aided in clinging when first seen is more likely to have been born earlier in the range of possible birthdates. In cases where the infant is clearly at least a few days old the Team might also note in the remarks or in demography or reproductive notes that the infant “looks old” or is “active” or is “breaking from the mother”.

In cases where there are no good clues, the midpoint between the day after the mother was last seen without the infant and the day she was first seen with the infant should be used, as when determining dates of death.

When observation is occurring at a normal frequency (every 1-4 days) then the infant will automatically be assigned a bstatus of 0, indicating that the birth is known within a few days. If there is a longer gap, clues such as umbilicus, blood on the mother, or notes indicating that the infant is very new can still be used to assign a birthdate with a bstatus of 0 since these things indicate the infant was newborn. If, however, the clues are less clear, then the infant may need to be assigned a bstatus of 0.5 or higher if the birth cannot reasonably be estimated to within a few days in either direction.

There are some rules of thumb that can help with assigning birthdates within a larger observation gap. Typically in these situations neonatal sheets are not available and there is usually little indication of the infant’s age other than that it is called a black infant (until up to about 4-6 months old) or a brown infant (about 4-6 months up until switched to a juvenile sometime in the second year of life).

If the conception date is closely known but the birth occurred during a large gap with none of the good clues noted above, the birth should be estimated by adding the median gestation time (178 days) to the conception date. The bstatus in such cases should be 0.5 since we know that the infant was born within a few weeks of the birth but not within a few days. A bstatus of 0.5 can be assigned for a birth known to within 3 months in either direction, though most such births are known more closely. If the date the infant is first seen is within 29 days of the resulting estimated birthdate, the infant should have an entrytype of B (like bstatus = 0 individuals do) since the infant will be interpolated present continuously between birth and the date first seen. If the date first seen is more than 29 days after birth, then an infant with bstatus 0.5 or higher should be assigned an entrytype of O. Earliestbirth and latestbirth should then be calculated based on all available demographic and reproductive cues available. Because infants with bstatus = 0.5 and entrytype B could legitimately have been born up to 29 days after their entrydate, latestbirth is permitted to be up to 29 days after entrydate for these entrytype B individuals. By contrast, all entrytype O and I, or left-censored, individuals must have latestbirth <= entrydate, which is why entrydates for these individuals were moved in Feb 2020 from the first date seen in members to the first date seen in census. Earliestbirth should not be more than 0.25 years prior to birth and latestbirth should not be more than 0.25 years after birth for individuals with a bstatus of 0.5.

If there is less data available then a bstatus of 1 or occasionally even 2 (in rare cases where we know there was an infant but we're not quite sure how old the infant was and whether or not an early fetal or infant loss preceded the birth of that infant such that there is a window of more than 6 months in one direction or the other from the estimated birthdate) may be appropriate. If the conception date is unknown or not well known but the birthdate is known to have occurred within a gap of a few weeks or months, as may occur in a very gappy period or in groups that have been recently dropped, the best we may be able to do is to estimate the birthdate using the midpoint between the date the mother was last seen without the infant and the date she was first seen with the infant. The conception date can then be backestimated to 178 days earlier, or if the conception date is also known to within a few weeks or so the midpoint might also be used for the conception date, provided that a realistic gestation time results (roughly 178 $\pm$ 20 days). As with estimating the conception date (see Appendix 7), there may also be intermittent clues that allow a more precise placement of the pregnancy such that all available reproductive data within the range of dates for the pregnancy should be considered when attempting to assign birthdates under these circumstances. In these cases, and those were an infant has a bstatus of 0.5 but there are more than 29 days between the estimated birthdate and the date first seen, the individuals should be assigned an entrytype of O. The entrydate must equal the date first marked present in census with a C or D census row. Earliestbirth and latestbirth should again be calculated by taking into account all available demographic and reproductive clues and any

descriptions of the infant or juvenile's development. The difference between birth and earliestbirth and that between birth and latestbirth can be no more than bstatus/2 years.

There are also a few guidelines for assigning birth estimates to fetal losses. Clues that a miscarriage is occurring or has recently occurred include blood on the mother's perineum, loss of pink colour on the PCS, and starting a new cycle.

Sometimes perineal bleeding is observed early in the pregnancy, around the time the female would normally menstruate but this does not appear to be indicative of an impending fetal loss. Occasionally bleeding later in the pregnancy also occurs with no apparent effect and a healthy infant is born. Other times bleeding is the result of a fetal loss. This is usually followed up quickly with loss of pink and/or beginning a new cycle but it may occasionally occur in fits. Thus the first sign of bleeding is not necessarily the best place to put the birth/death date. In these uncommon cases where bleeding is observed more than once over a long period hormone data might be helpful in narrowing down the actual date of fetal loss. But typically the first sign of bleeding is soon followed by loss of pink and/or the start of a new cycle, in which case that first date when bleeding is observed is assigned as the birth and statdate for the fetus.

In other cases no perineal bleeding is observed and the first sign of fetal loss is loss of pink on the mother's PCS or the mother starting a new cycle. Loss of pink is almost always indicative of fetal loss so (unless an infant shows up anyway) a fetal loss should be assigned when loss of pink is first noted. A female beginning a new cycle without having given birth also usually indicates a fetal loss and the birth/death date of the infant is assigned as the female's tdate. Sometimes, however, a female will have what the Team refers to as a "funny swelling" where she has a swelling (typically but not always small and brief) while she is pregnant. This usually occurs around the time a female would typically begin her next cycle were she not pregnant. If, however, she goes on to start showing pink and other indications (records of consortships, timing of pink showing up on her PCS, gestation time) suggest that she conceived before this funny swelling then the funny cycle is ignored.