

ABRP Data Management Protocols: Behavioral and Group Movement Data

Jacob B Gordon

30 Nov 2022

*Based on previous versions written by
Susan Alberts, Leah Gerber, Lacey Maryott,
and Patrick Onyango*

Table of Contents

General Introduction	4
Filing and Management of Amboseli Data	6
Filing Original Data	6
Reviewing Original Data.....	7
Filing Photocopied Data	8
Electronic Data	8
Babase Database	11
Basics for Data Entry and Management	11
Log Books	11
Rules for Data Entry	12
Naming Conventions for Representative Interaction Data	12
Data Management	13
Compares	17
REPRESENTATIVE INTERACTION DATA	18
Changes to Scanned Data	18
Data Entry	19
Agonisms/Groomings (“Ad-Libs”)	19
Mount and Consorts	23
Multiparty Interactions (“MPI’s”)	25
Validation and Update	29
Checking babase for interaction errors	31
Comparing interaction data with census records	31
FOCAL SAMPLING DATA	33
Proofing	33
Validation and Update	34
Common Errors.....	34
GPS/SWERB DATA	36
Processing data from the field	36
General guidelines for proofing GPS data:.....	38
Correcting Errors and Special Cases in GPS Data	41
Groves and Waterholes	43
Other SWERB/GPS rules	44
Data Management in Babase	44
Data Upload	44

Deleting/Updating Data	44
SWERB_LOC_DATA_CONFIDENCES	44
DARTING DATA	46
Data Storage	46
Data Entry	46
Data upload	47
PREDATION AND HUMAN DISTURBANCES.....	48
QuickTapSurvey Management	48
Survey Upkeep	48
Data Management and Archival	49
Babase Data Management	50
The Predation and Human Disturbance Tables	50
Entry Codes for Predation Tables	55
Data Entry	56
Proofing Data	56
Upload Data	56
RANKS.....	57
Updating Ranks	57
Decision Rules for Assigning Male Ranks.....	58
Ranked “On” and “By” Dates	61
Wiki and Other Online Management Tasks	65
Updating Dataset Status Table	65
System Email Management	65
Others	65
Appendices	66
Appendix 1: Excel Dates (Mac versus PC)	66
Appendix 2: PLHD Project, Creating tables	66
Biography	68
Fertility	68
Appendix 3: Feedback to Field Team	69
A: Rate of data collection.....	69
B: Observer Error	71

General Introduction

This document contains all the necessary procedures for management of the Amboseli Baboon Research Project's behavioral and group movement data, including descriptions of procedures for handling incoming data from the field (electronic and paper) and detailed guidelines on how to enter and update information in the database. As of this writing, this management is performed at Duke University, so this and other ABRP documents may refer to the "Duke" data rather than the specific data sets.

Protocols for managing demographic and reproductive data are contained in a separate document, available on the [Babase Wiki](#). That management is performed at Princeton University, so ABRP documents may refer to the "Princeton" data rather than those specific data sets.

A few basic rules:

1. Always document/record/back up your work
2. Keep up with the data entry and proofing
3. Keep track of all data entry, proofing, and updating in the appropriate log books
4. Keep track of emails from Princeton to field team. These contain important information on updates of various data types.
5. Enter Amboseli data in the appropriate subdirectories of the "ABRP DATA MANAGEMENT" folder in our shared Dropbox.
6. Document any new protocols or protocol changes

Most of the data management should follow these steps:

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

The following sections of this document explain in detail how these steps are followed with the different types of data. This manual is only for managing the data entry and updating. In depth details about the Babase data management system and detailed descriptions about the data and data collection can be found in the *Amboseli Baboon Project: Data Management System, Pocket Reference*, and the *Guidebook for the Long-term Monitoring of Amboseli Baboons and their Habitat*, respectively. These guides can be found online at the following url's:

https://papiro.biology.duke.edu/babase_system.html

https://papiro.biology.duke.edu/pocket_reference.pdf

<http://amboselibaboons.nd.edu/downloads/> (click on "Monitoring Guide")

Filing and Management of Amboseli Data

Original data is kept in the Alberts lab library at Duke. Original data remains in Amboseli until someone traveling from the field brings it to the US. Some of the data are scanned in the field at the end of each day and emailed to project leaders and data managers. At the end of each month, all the remaining (not yet scanned) data from the month are scanned. All of these scans are saved in Dropbox folders to make them available to project leaders and data managers.

Filing Original Data

Monitoring Data - There is a yearly monitoring binder for each study group. The following types of data are filed in the monitoring notebooks: census and sexskins, demography notes, reproductive notes, subgroups, daily data (including agonisms & groomings a.k.a “ad-libs”, mounts/consorts, and multiparty interactions), other groups, wounds and pathologies, and other notes.

Older binders may be different, as follows:

- Before May 2018, there will be separate sections for groomings, decided agonisms, undecided agonisms, mounts/consorts, and multiparty interactions. In that month we changed the way that observers record those data, such that it no longer made sense to separate them.
- Data about predations, predator sightings, and human disturbances were written on paper sheets until late 2017, when we switched to an all-digital form. Notebooks through 2017 will have “Predations/Human Disturbances” sections containing those data.

Principles to remember when filing new original data:

1. Use the existing binders as a model.
2. Make sections exactly as in previous binders.
3. Census, Demography, Sex Skins, and Reproductive Notes go together.
4. Check to make sure all the pages are there. They should be numbered in the field by the team—“1 of 5”, “2 of 5”, etc.—but also be aware that sometimes they are numbered incorrectly. So use the numbers as guides, but don’t take them as gospel.
5. Note big gaps in data that you notice, e.g. missing 6 months, or an entire group.
6. Try to fit a whole year in a single binder.
7. Tape the bottom of each sleeve so the data stays in place but is still accessible from the top

Meteorological Data – Data on ambient temperature and rain gauge readings at the camp. Filed in chronological order in the “Meteorological Data” notebook.

Charcoal Fridge Temperature – Max-min data for the temperature of the charcoal fridge. Filed in the “Meteorological Data” notebook.

Half-hourly Temperatures – Half-hourly weather recordings are now sent as electronic files so don’t be expecting to add any more data to this section of the “Meteorological Data” notebook.

Neonatal Assessments – There is a notebook called “Neonatal assessments” Both neonatal assessments for each new individual get filed in here. Neonatals are filed by group, alphabetically by the mother’s name, and then infants in birth order for each mother.

Canine Condition and Scrotal Development – These two types of data are filed in the same notebook: “Canine Condition and Male Age Estimates” by data type, group, and year.

Hybridity Scoring – filed in the “Hybrid Scoring” binder.

Male Age Estimates – filed in the “Canine Condition and Male Age Estimates” binder.

Other Group Censuses – When possible, census data are collected for several other baboon groups in the Amboseli region including Joy’s, Nzige’s, Proton’s and Stud’s. These data are filed in the “Other Group Censuses” notebook chronologically by group.

Monthly Summaries and Satellite Phone Logs - File in “Monthly Summaries” box.

Daily Activity Calendars - File chronologically in the “Activity Calendar” binder by person.

Fecal lists – Lists of fecal samples are put in the “Fecal Lists” box

Tree Grove Monitoring – These data are filed in chronological order in the “Tree Grove Monitoring” notebook. [Note: as of 2011 these data are no longer collected.]

Reviewing Original Data

When filing original data try to get a feel for what is going on and make a note of any missing periods of data that we will need to track down. You may need to

check email correspondence from the previous year (e.g. a clarification may have come in that XXXXX's group only had 4 pages of agonisms, not 5). Or check our digital copies of the monthly scans.

Filing Photocopied Data

This whole section is no longer accurate. Since 2020, entry of new data is done using PDFs of the scanned data, not actual paper printouts of them. We retain this section only for historical purposes: to indicate how data were previously entered.

Photocopied data gets sent from the field each month to Nairobi, where it is scanned and emailed to us. Most of those data can simply be archived electronically (see Electronic Data, below), but some sheets need to be printed off for data entry. The following data should be pulled out of each data packet:

1. Census and Demography notes
2. Body fat data (if any)
3. Darting logs/sheets

Representative interactions and census/demog notes are filed in the "Working Photocopies" binders organized by group. Each of the other data types have their own respective binders. When the data come in follow these steps:

1. Check to make sure all the data is there (no missing groups or pages).
2. Check to see if the photocopies are clear and no edges have been cut off
3. Read the demography notes and check for any new grove and waterhole information (see GPS data section)
4. Mark any data changes on the photocopies. These changes will usually come by email from the field team, or the Princeton data manager. Princeton checks the data thoroughly and clarifies any questions with the field team. The questions and responses get cc'ed to Duke. These responses likely won't come until after you've already filed the data, so you'll need to back to those data and make any necessary corrections. (In years past, the data came to Princeton first, and you might have corrections in your email box before you received physical copies from Princeton.)

Electronic Data

All electronic data are saved in the Alberts Lab's shared Dropbox, as subdirectories in the **Alberts Lab/ABRP_Data Management** directory. Each observer in the Amboseli field team emails electronic data files daily, or sometimes every other day. These emails contain three types of data:

- Focal sample data: CSV files exported from AmboPrim8
- GPS readings: TXT files exported from MapSource

- Scans (PDFs) from their field notebooks

The “daily” emails on Saturdays will also include some extra files that are sent weekly:

- Weather data: DAT files exported from Visual WeatherHawk

Princeton handles the weather data. The rest is for the Duke manager. When a new email comes in, unzip the attached files and save the files to the appropriate year-month/data type folder in **ABRP_Data Management/Data from Amboseli**. The GPS and focal sample data should also be copied into appropriately-named ‘YYYY-MM’ folders within their respective ‘in progress’ folders: **ABRP_Data Management/DATA/GPS/In Progress** or **ABRP_Data Management/DATA/FOCAL SAMPLES/In Progress**.

As mentioned previously, at the end of each month all the handwritten data—except for the field notebook pages that were already scanned and sent—are scanned in the field and saved in a Dropbox folder. You should copy these “monthly data” and save them in the appropriate year-month/“Monthly Data” directory in **ABRP_Data Management/Data from Amboseli**. When all of the month’s data have been received, prepare them for data entry by doing the following:

1. Make a new folder to contain all of the month’s “Working Data”. There is a template in the [Dropbox]/Alberts Lab/ABRP_Working Data/Monthly Handwritten Data, named “mm_Mmm 202X”
 - a. E.g. the data for December 2020 would be “12_Dec 2020”. Using the number in the name makes sure that monthly data are sorted chronologically.
2. In this working folder, add an Excel sheet that your data entry techs can use to log their work. Again, there’s a template in the aforementioned template directory.
3. Within this working folder, make a new folder called “Entered copies”. This is where your data entry techs will save their copies of the data.
4. Also within the working folder, make a new folder called “Handwritten Copies”, containing separate folders for each group.
5. Save new PDFs with the groups’ census and demography pages, one file for each group. Each file should go in its respective group’s folder in “Handwritten Copies”.
 - a. To keep this file distinct from the other files you will store in this folder, it is suggested you give the file a name that starts with the number “1”, e.g. “1 ACA DEC 20 CenDemog” for Acacia’s data in December 2020. It doesn’t really matter, though.

6. Save new PDFs of each of the PDFs of “daily” data (ad-libs, mounts/consorts, multiparty interactions, focal sample records), renaming the files so that they’re named by observer, then the date, then the group.
 - a. E.g. a file named “20201225 JKW ACA” would be renamed “JKW 20201225 ACA”.
 - b. Renaming every file is kind of a lot of tedious work if you do it manually. You’ll probably want to use a script to do it for you.
7. Store each of the above “daily data” copies in its respective group’s folder. If you correctly renamed all the files, the files will be sorted in the order in which they should be entered.
8. Move this working data folder to a Dropbox directory that you share with your data entry techs, so they can see the PDFs and enter their copies.

Later in the month, you will also receive a few other electronic files that should be saved in the same **.../Data from Amboseli/YYYY_DD/Monthly Data** directory. Specifically:

- Agonisms and Groomings (Excel sheets), entered by SNS and BOO
- Fecal Supplies Inventory (Excel sheet), from BOO
- The prim8 individuals import file (CSV), from BOO
- Agonism matrices (Excel sheets), from SNS
- The monthly report (RTF)

The prim8 individuals import file is usually named simply “prim8_import_individuals.csv”. When you receive it and save it in this directory, rename it to indicate the day you received it from the field, e.g. “prim8_import_individuals_recd_20170403”. This “received” date can be helpful information if problems or discrepancies arise in the focal sampling data.

A copy of the agonism and grooming files should also be saved as TXT files, named as the “B” copy according to the naming conventions for those data (see Naming Conventions for Representative Interaction Data below), and saved in the appropriate data type folder in **ABRP_Working Data**. You should also add column headings to the file, and rearrange the column order to match you’re your data enterers will use (see instructions for data entry). When finished, you log the existence of these “B copies” in the appropriate data entry logbook (see Log Books, and Naming Conventions for Representative Interaction Data).

Babase Database

Babase is a PostgreSQL database, written initially by Karl Pinc and later by Jake Gordon. For technical information about Babase, see the tech specs at http://papiro.biology.duke.edu/babase_system_html/. The source code can be viewed at <http://papiro.biology.duke.edu/babasewiki/SourceCode>.

To access Babase, go to <https://papiro.biology.duke.edu/phpPgAdmin/>. A login can be requested through the data managers (including you) or Karl Pinc. You should also be given an additional “admin” account for some of your occasional needs as a data manager.

Information about how up to date the various datasets are can be found on the Babase wiki, <https://papiro.biology.duke.edu/babasewiki/DatasetStatusTable>. As a data manager, you should update this page regularly whenever you update a dataset. The page does not update itself.

Basics for Data Entry and Management

Most data entry should be performed by workers, usually undergraduate work study students. You are the data manager, not the data enterer. However, you should enter every type of data at least once. This ensures that you’re familiar with how it is recorded, common mistakes that observers can make, and common mistakes that enterers can make.

Many different types of data are being entered and utilized for the Amboseli Baboon Research project. Each of these types of data has a unique format and several programs are used in the process of entry, proofing, validating, and updating. Nearly all data types used in the project adhere to the following system for management.

Log Books

Each data type has its own log book containing a record of the current status of the data and all procedures as they are completed. These procedures include: arrival (electronic data), entry, comparison, validation, updating, and archiving of the data. Once each of these procedures is finished, it is checked off in the log book. Thus, the last check made reveals the current status of the data set.

Until 2020, there were actual paper log books used to track all this information. When the COVID-19 pandemic forced everyone to work remotely, we switched to digital logbooks, i.e. Excel sheets.

Rules for Data Entry

For representative interaction sampling (“RIS”) data, two datasets are entered independently by two different people. This allows for proofing by comparing the two datasets against each other. The data sheets must be entered in exactly the same order for this proofing process to work. In general, start data entry at the top left of the page and proceed to the bottom right, entering data in each column in the order it was hand-written. For extra insurance, the person entering the “A” copy could note the order of entry on the scanned data pages so that the “B” copy will match. Numbering and good notes during the data entry phase can save a lot of time and confusion later on.

On occasion, letters or numbers are omitted or unclear on the scans. It is possible that the originals may be discernible so omit the line for now and check the originals when they arrive. Clearly note the problems by adding a note to the PDF. Ideally, entry technicians will alert the data manager about these with an email, Slack DM, etc. If the question is related to something recent, the data manager can email the observer and ask for clarification. If the observer cannot resolve the problem, or it’s too long ago for them to plausibly remember the answer, do not enter the record and make a note on the scan.

Naming Conventions for Representative Interaction Data

All file names follow the same general naming convention. Files should be saved in tab-delimited text format (.txt), and named as follows:

G T Y Y M M A/B .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 file. The correct letter for each group is available in Babase, in the GROUPS.One_Letter_Code column.

T = Type of data. The types of data that are entered are **AD**-Libs (Groomings and Agonisms), Mounts/**C**onsorts, and Multi**P**arty Interactions. Before May 2018, **G**roomings and **A**gonisms were entered separately.

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.

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 etc.

A/B (A or B) = The proofing system requires two sets of data to be entered. Use **A** for the first set entered and **B** for the second set entered.

Historical note: sometime in the early 2000's, this standard format was changed to use the month before the year, MMY. Beginning with the January 2022 data, the standard was reverted to using YYMM.

Data Management

As data are entered, proofed, and archived, the files will be moved between different directories. The series of directories is used mostly to indicate which stage of management a particular file needs, but is also used to control users' access to the files.

Data Entry: when workers first start entering data, they need a shared space to save their work that is accessible to you. The “Data Entry” directory in **Dropbox/Alberts Lab/** serves this purpose. Your workers will need Dropbox accounts to access the folder; when they make those accounts, share this folder with them.

We use this small, dedicated directory for data entry partially for security reasons, but also out of concern for available space in an undergraduate's Dropbox account. The full “Alberts Lab” directory contains hundreds of GB of data, which far exceeds the capacity of most undergrads' usually-free Dropbox accounts.

Working Data: after workers have finished entering a sheet of data, you should move it to a private space. In **Alberts Lab/ABRP_Working Data/**, there are various subdirectories, each named according to the type of data it contains (“AGONISMS”, “GROOMINGS”, etc.). Roughly once a week, you should move completed files from the “Data Entry” folder into the appropriate subdirectory in ABRP_Working Data. If a file is stored here, it means that it has been entered, and awaits “compares” (see below) between the A and B copies.

It could be argued that this “Working Data” directory is not needed. Before we moved all our data to Dropbox, Working Data was where workers saved their entered data AND where files awaited proofing. With the migration of our data to Dropbox, the separate “Data Entry” folder had to be made, so it was shoehorned in to our usual process. For now, we have retained the “Working Data” directory in this process, simply because “if ain't broke...”. Future Duke managers may want to consider removing it and going straight from “Data Entry” to the “In Progress” folder.

In Progress—After compares are complete, move the files (“A” copies only; the “B” copies should be deleted) to the appropriate subdirectory in **ABRP_Data Management/DATA/[data type]/In Progress**. These data are almost ready to get into Babase, and wait here until it's time to import the data.

These data are not yet fully proofed; there still may be some errors that will not be detected until import to Babase (see

Multiparty Interactions (“MPI’s”)

All of the data discussed so far has been about the various kinds of interactions that occur in dyads (between two individuals). Multiparty interactions are just what their name suggests: interactions that involve more than two individuals.

Entry of MPI data is more complicated than dyadic interaction data. You should not enter MPI data until you have experience and confidence with the dyadic data.

Although multiple parties are involved, MPI’s are recorded by observers as a series of dyadic interactions. See the example below. This MPI from RSM in Narasha’s group says:

- NOI AS’d LYQ
- LYQ asked an unknown individual* for help
- LYA helped LYQ
- LYA groomed LYQ

*In this context, “unknown” may mean one of a few things. LYQ may have screamed for help from anyone rather than a specific individual. Or he may have intended a specific individual, but because our observers lack the ability to read minds they could not identify which individual was the intended recipient. “Unknown” does not mean that the observer simply doesn’t know the name of the requested individual.

Let’s talk about how these data should be entered. First, the column headings.

Column Name	Description
mid	A number that identifies the row within the uploaded file. The number must increase with each row and must be sequential.
coal_id	This number is an identifier which associates all lines within a single interaction. All rows associated with a given MPI must share the same number.
grp	Contains the group number where the animal was on the day of the interaction. See below for more specific information.
date	As above, use the format dd/mm/yyyy
actor	This column can contain more than one actor in instances where the interaction is not distinctly dyadic. If more than one actor, use a comma—and no spaces—to separate them.
agg_act	This is the “act”. See below for more information.

recip	This is the “actee”. It can contain more than one actee in instances where the interaction is not distinctly dyadic. If more than one actee, use a comma—and no spaces—to separate them. See below for more information.
outcome	Contains responses for those rows with an act of “?”, and should be empty otherwise. See below for more specific information.
form_passive_aid	Leave blank. This column doesn't have a purpose, yet.
context	Indicates the context in which the interaction took place, if the observers noted anything unusual. This value should only be provided on the first line of the MPI; it must be blank on all other lines. When context_type (below) is “C”, this column must be “CONSORT”.
consort	Indicates the outcome of the MPI when the “context” is “CONSORT”. See below for more-specific information.
context_type	Contains a code for the “type” of context in which this multiparty interaction took place. Each MPI must have a context_type on its first line. When a consort is involved, the context_type is “C”. Otherwise, this should be “N”.

You may notice the absence of an “observer” column. Perhaps—hopefully—one day we’ll start recording the observer, but for now, we don’t.

The “Grp” column

Our database uses unique numbers for each group. So instead of the group’s name, you need to provide the group’s number. Below is a list of all the current study groups’ numbers.

Acacia’s: 1.211

Hokey’s: 1.11

Narasha’s: 2.11

Europa’s: 1.2222

Yoda’s: 1.2221

Trailing zeroes are okay to add, but also not necessary. For example, 1.11, 1.110, and 1.1100 are all acceptable ways to refer to Hokey’s group.

The “Agg_Act” column

MPI’s may use any of the acts used for agonisms and groomings. There are several other agonism codes that are exclusive to MPI’s that you may see. These are too numerous to include here, but they’re fairly similar to the ones you already know. If you see a code you don’t recognize, enter it just as the observer entered it, or ask the data manager.

There also a few new acts that are not agonism-related and occur frequently in MPI’s. These acts are:

- ? Actor asked actee for help
- + Actor helped actee
- P Actor gives **P**assive help to actee

The “Recip” column

In lines where help is requested (lines with act “?”), the actee is often unknown, or “UNK”. *In these cases, enter “U”, not “UNK”.*

The “Outcome” column

In the “?” rows, the observer may write a short code after the actee, indicating how the actee responded to the request for help. When recording these “outcomes” of help requests, you will enter something slightly different from what the observer actually wrote. See below.

Outcomes

Code	Meaning	How it is recorded
+	Actee provided help	SUCC
-	Actee did not provide help	FAIL
P	Actee provided passive help	PASS

When the actee is “UNK”—which should only ever happen in a “?” row—the “outcome” must be entered as “FAIL”.

The “Consort” column

When the observer notes that the MPI was part of a consort, s/he will include comments indicating who was consorting whom, and whether he kept the consort or if another male took over the consort. Those data are recorded in this column using a very specific format. Any data in this column should always be of the format:

ADM WITH ADF;ADM GET ADF
 or
 ADM WITH ADF;ADM KEEP ADF
 (No space after the semicolon!)

Any deviation from this structure, such as spaces, or different wording, will create an error.

Examples

See the below MPI sheet. It’s a bit old, from a time when MPIs were recorded on their own sheets instead of sharing a page with Mounts & Consorts, but otherwise nothing about the data has changed.

These data would be entered the data as shown below.

mid	coal_id	grp	date	actor	agg_act	recip	outcome	form_passive_aid	context	consort	context_type
1	1	1.211	07/03/2018	CAX	AS	ACC			CONSORT	CAX WITH UNA;CAX KEEP UNA	C
2	1	1.211	07/03/2018	YOI	P	ACC					
3	1	1.211	07/03/2018	OBL	P	ACC					
4	2	1.211	30/03/2018	EID	AS	IGL					N
5	2	1.211	30/03/2018	ERM	+	EID					
6	3	1.211	30/03/2018	ACA	AS	OPA					N
7	3	1.211	30/03/2018	ACI	+	ACA					
8	3	1.211	30/03/2018	APP	+	ACA					
9	3	1.211	30/03/2018	AKP	+	ACA					
10	3	1.211	30/03/2018	ERM	+	ACA					
11	3	1.211	30/03/2018	EID	+	ACA					
12	3	1.211	30/03/2018	OBL	+	OPA					

Validation and Update).

Final Data—After you’ve successfully imported data to Babase, the final copies of the data are stored in the “Final Data” folder for that particular data type (usually by year/month), in **ABRP_Data Management/DATA/[data type]/Final Data**.

Compares

The A and B copies should be compared to each other, in order to screen out typos and misread handwriting. A simple web client to compare two tab-delimited text files can be found on the Babase wiki: <https://papiro.biology.duke.edu/programs/wwwdiff/>. When discrepancies are found between the two copies, you will need to resolve them. Check the handwritten sheets for the “right” answer, then correct the files as needed. Continue resolving discrepancies until the files are identical. When this is true, you no longer need both copies and can delete the B copy.

REPRESENTATIVE INTERACTION DATA

The interaction data are entered monthly, by group and by data type. Each group has a separate file for each of the three classes of interaction:

- (1) Agonisms/Groomings (i.e. “Ad-Libs”)
- (2) Mounts/Consorts/Ejaculations
- (3) Multiparty Interactions

Changes to Scanned Data

When there are errors or omissions in the data corrections or clarifications need to be noted. Once or twice each week, you should go through the recently-sent data and check for any obvious errors. You won't be able to recognize every possible error, but there are several that you can recognize, including:

- Missing data:
 - Observer initials *on each sheet*
 - Group name *on each sheet*
 - Date *on each sheet*
 - Whole sheets missing (a mounts/consorts and MPIs sheet should be sent, even when there's no data)
 - End times for consorts
 - Actor or actee (in ad-libs)
 - Agonisms are in multiparty interactions but not also in the ad-libs
 - When multiple observers are at the same group, one of them needs to identify himself as the “primary” observer of the mounts/consorts, and the other(s) need to identify as “secondary” or “not primary”.
- Mistakes
 - Wrong date, on the individual sheets and/or the file name
 - It's not unusual to be off by a day, e.g. write “16” when “17” is the correct date
 - Sometimes they write the wrong year, especially in January and early February
 - Occasionally they write the wrong month, especially at the beginning of a new month
 - Wrong group, on the individual sheets and/or the file name (not as rare as you might expect)
 - Implausible mounts/consorts
 - While a female is being consorted, it is unlikely that any other male will mount her. When there are multiple consorts, sometimes the observer mixes up who was with who, e.g. if A is consorting B and C is consorting D, we might see A mounted D.

- Wrong individuals. This is pretty much impossible to recognize unless you're very familiar with the population and who is in which group. Even with that familiarity, it's difficult. Still, you may notice some of these mistakes, including:
 - Misspellings, e.g. "EXB" for the individual "Xbox", or "EQU" for "Ecuador".
 - ID switches, e.g. if two infants are born at about the same time in different groups, the observer might mistakenly use the other group's infant name to refer to "this" group's infant. (This is rare.) Likewise for immigrants who appeared at around the same time. (Also rare, but more common with immigrants than infants.)

Keep a list of errors like these that you might find. Once each week, send an email to the field team and ask them about these. It has been our experience that one weekly email with questions for the week tends to work better for the team than if we send several single-question emails throughout the week.

The Princeton data manager checks the daily data too, for these and other issues. Both of you are imperfect and may miss something, but between the two of you, few mistakes slip by. The Princeton manager also goes through the monthly monitoring data and checks for other issues. They will compile their questions into an email for the team, and you will be included in that conversation. If there are changes to the data, they should be made in the relevant PDF files. It is preferable to do this before the data gets entered. If it isn't, it's important that you take care to update the entered copies appropriately as well.

Data Entry

Rules for data entry vary slightly for the different types of interaction data. In general, two copies of each dataset are entered, to allow for proofing later. Because the proofing program works by comparing the datasets line by line, entering both versions in the same order is critical. In addition, data enterers should clearly note any observed errors directly in the PDF files.

The remaining text in this section is copied from the data entry protocol, hence the occasional tendency to refer to "you" and to direct the reader to consult the data manager.

Agonisms/Groomings ("Ad-Libs")

These data are recorded by the observers using the general format:

ACTOR BEHAVIOR RECIPIENT

The actor and recipient—or “actee”—are recorded with a 3-letter abbreviation for each individual’s full name. For example, ‘Acid’ is abbreviated “ACI”. The abbreviation is usually but not always the first three letters of the individual’s full name; ‘Noggin’, for example, is abbreviated “NGG”.

The behavior—or “act”—is also abbreviated with a one- or two-letter abbreviation. These codes are:

- G Groom. Actor **G**roomed actee.
- AS (Agonism) Actor acted **A**ggressively, actee **S**ubmitted
- OS (Agonism) Actor not aggressive (**O**ther), actee **S**ubmitted
- dS (Agonism) Actor **d**isplaced actee, **S**ubmission implicit

An example of an “Ad-Lib” data sheet is shown below. A discussion of how to enter it follows.

JKW YODAS 01 JUN 18 2/3

GROOMINGS AND AGONISMS

YAP OS VUR	ADL AS YOD	VLA AS BEX
KOM G RAN	KOM AS RAN	EAS G EXE
EAS G NER	KOR OS ROB	KOM OS ADL
KOM G EAS	NER OS VAI	*TRU AS KOM
ROD OS RHO	KOR AS ROB	*TRU AS ADL
RUI OS YEL	ADL AS KOR	*TRU AS LUZ
TRU OS RWA	ROD G REN	*TRU AS KOR
TRU OS DIW	ROD G RAN	
DIW OS RHO	KOM AS RAN	
ROD OS RUE	EAR AS YOD	
DIW OS RHO	*RAK AS YEL	
ROD G RAN	*RAN AS YEL	
RUI G ROB	YEL OS VUR	
ADL AS EAS	RWA OS RAZ	
DIW OS RAK	VLA OS R	
DIW AS ADL	RAZ AS RAZ	

Entering the data

When entering these data in Excel, the data will look like this:

date	actor	act	actee	observer
01/06/2018	YAP	OS	VUR	JKW
01/06/2018	KOM	G	RAN	JKW
01/06/2018	EAS	G	NER	JKW
01/06/2018	KOM	G	EAS	JKW

(and so on, until...)				
01/06/2018	TRU	AS	LUZ	JKW
01/06/2018	TRU	AS	KOR	JKW

A few important points to emphasize:

1. There are five columns that need to be entered: date, actor, act, actee, and observer. These headers must always be entered in lowercase.
2. Everything else you enter should be in uppercase, regardless of the case used by the observer (e.g. “dS” should always be entered as “DS”, “g” should always be “G”).
3. Actor and actee names should always be three letters. If you see a name with more or less letters than this, check with the data manager.
4. The observer is a three-letter set of initials, usually written in one of the upper corners of a sheet. Most of the data is collected by one of the following individuals:

RSM: ~1981-present
 SNS: ~1989-present
 JKW: 1995-present (the observer in this example)
 ILS: ~2012-present
 SCA: 1984-present (but mostly stopped after ~2000)

Sometimes the observer forgets to write their initials on a sheet. With time, you will learn to recognize their handwriting, but the data manager can also show you other ways to determine the observer. Don't hesitate to ask the data manager if you're at all unsure.

5. The pages may be numbered (e.g. the “2/3”), but you should mostly ignore this. That is, if the order in which they appear does not match the numerical order of the pages, you should still enter the data in the order that it appears on the copies. Also, it's normal for some pages to appear to be missing, because some of the original numbered pages do not include agonisms and groomings.
6. Sometimes a behavior is recorded with an asterisk (*) next to it. Enter these interactions like normal, omitting the asterisk. The meaning of these is discussed later, under Multiparty Interactions.

Issues to look out for

When entering these data a few problems may arise. Please keep a look out for the following situations:

1. **Self grooming/agonism (such as OFR g OFR)**

Circle these entries on the scans and write “not entered” or “DNE” (“did not enter”) with the date and your initials.

2. **“White out” with nothing entered**

Circle these entries on the scans and add the date and your initials. Do still enter them, but leave the whited-out field blank or insert a “???” (or something like that) in its place. The data manager can contact the observers and ask for clarification.

3. **Names that are unclear (such as DOU and DOV).**

If a name is unclear or not 3 letters, go back to the census sheets to see which individuals were actually present in the group that day. An example census sheet is shown at right.

On the census sheet an “X” is marked every day the individual was present. The groups were observed on days marked. If an entire column is blank the group was not observed that day. When an “O” is present the individual was missing that particular day.

If DOU was present and DOV was not then you can enter DOU. If neither or both were present then we cannot determine which individual was intended and the row cannot be entered. When this occurs, err on the side of caution and don't enter the data. No guessing! Circle it, add the date and your initials, and a few words explaining the situation, e.g. “DNE, unclear if DOU or DOV”.

1 of 2
LINDA'S JUN 05

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
ARS	X		X		X	X					X	X		X	X		X		X	X	X										X
EDU	X		X		X	X					X	X		X	X		X		X	X	X										X
EXO	X		X		X	X					X	X		X	X		X		X	X	X										X
HIB	X		X		X	X					X	X		X	X		X		X	X	X										X
TDI	X		X		X	X					X	X		X	X		X		X	X	X										X
KRA	X		X		X	X					X	X		X	X		X		X	X	X										X
MOR	X		X		X	X					X	X		X	X		X		X	X	X										X
PLA	X		X		X	X					X	X		X	X		O		X	X	X										X
SAW	X		X		X	X					X	X		X	X		X		O	O	X										X
TIN	X		X		X	X					X	X		X	X		O		O	O											O
TOG	X		X		X	X					X	X		X	X		X		X	X	X										X
WIV	X		X		X	X					X	X		X	X		X		X	X	X										X
WAB	X		X		X	X					X	X		X	X		X		X	X	X										X
LEX	X		X		X	X					X	X		X	X		X		X	X	X										X
LAW	X		X		X	X					X	X		X	X		X		X	X	X										X
WOO	X		X		X	X					X	X		X	X		X		X	X	X										X
WYL	X		X		X	X					X	X		X	X		X		X	X	X										X
NOZ	X		X		X	X					X	X		X	X		X		X	X	X										X
LES	X		X		X	X					X	X		X	X		X		X	X	X										X
NOK	X		X		X	X					X	X		X	X		X		X	X	X										X
NIR	X		X		X	X					X	X		X	X		X		X	X	X										X
WYC	X		X		X	X					X	X		X	X		X		X	X	X										X
NOJ	X		X		X	X					X	X		X	X		X		X	X	X										X
NAG																															X

Please remember that you can consult with the data manager or with Susan to resolve problems with names; remember also that because these names are hand-written, it takes a little getting used to!

If the writing is unclear but you are SURE which individual it is, circle the entry and write down what you actually did enter, including your initials and the date.

An example of an unclear entry and corresponding note is shown below. This is from an older page, when groomings and agonisms were separated, but the idea is the same.

JKW - LINDA'S GROOMINGS APR 05

11 APR 05	WON g	WAS	EXO g	WIP.
NOZ g	WIF	WAS g	WHA	EXO g
WIF g	NOZ	NAP g	LOL	LIW g
MBE g	KRA	LOL g	NAP	LIW g
KRA g	MBE	WRA g	NOB	WAS g
LIW g	W.	LAO g	TIN	LAO g
NOB g	NIR	MBE g	KRA	LAN g
NOB g	NOD	16 APR 05	LOL	g
NOD g	NOB	LIW g	LAW	NOB g
14 APR 05	LAW g	LIW	LAN g	LAD
NOB g	KAT	WAS g	WON	WAS g
NOB g	KAT	WON g	WAS	NOB g

Row not entered LG 13/12/05

If no observer initials are indicated, you may want to check the census record; the above sheet has data from the 11th, 14th, and 16th of April 2005, so if the census record shows that JKW was the only observer present on all three of those days, then you can infer that this sheet must be from him.

Mount and Consorts

Mounts and consorts are similar to the two previous types of data, with a few small differences:

1. These interactions use the following "act" codes:
 - E (Mount) Actor mounted actee, **E**jaculate seen
 - M (Mount) Actor **M**ounted actee, ejaculate not seen
 - C Consort. Actor **C**onsorted actee
2. Times are recorded for these interactions. We enter these in two new columns: "start" and "stop".
 - a. Mounts are recorded with a single time, which should be recorded in the "start" column. Leave "stop" empty.

- b. Consorts are recorded with beginning time (“start”) and end time (“stop”).
3. The sheet that has the Mounts and Consorts data may also have other kinds of data written, probably Multiparty Interactions/”MPI’s” and Undecided Agonisms. Those data are entered elsewhere, so don’t attempt to include them here.

When a consort time has an “(E)” written after it this is an artificial end time, likely the time that the observers ended the day’s observation on the group. We enter the time as is and ignore the “(E)”. You may also see a “(B)” to indicate an artificial begin time. Handle these the same way you handle “(E)”s: ignore them.

All times must be in 24-hr time (which is how they were recorded), and must be entered with a colon, eg: 09:15 or 14:15, not 0915 or 1415.

If there is no start or end time the row can still be entered, **leaving the related time field(s) blank.**

When a name is unclear it can be very helpful to look at other mounts/consorts for the same day. If a mount is unclear but there is a consort at the same time on the same day, we may be able to deduce who was involved in the mount.

Usually, mounts only occur between individuals who are also in a consortship. Occasionally this is not so, and the observers usually add an asterisk or something like it to emphasize it. When this happens, you needn’t treat the data any differently. Don’t enter the asterisk anywhere; the very fact that we have no consorts for that pair is evidence enough that it was outside of a consortship.

An example of a Mounts & Consorts data sheet, and the corresponding data entered are shown below.

Entry of MPI data is more complicated than dyadic interaction data. You should not enter MPI data until you have experience and confidence with the dyadic data.

Although multiple parties are involved, MPI's are recorded by observers as a series of dyadic interactions. See the example below. This MPI from RSM in Narasha's group says:

- NOI AS'd LYQ
- LYQ asked an unknown individual* for help
- LYA helped LYQ
- LYA groomed LYQ

NARASHA'S 02 JUN 18 RSM	
MULTIPARTY INTER.	MOUNTS/CONSORT
NOI AS LYQ	"SENS RECORDED"
LYQ ? UNK	MOUNT / CONSORTS
LYA + LYQ	
LYA g LYQ	

*In this context, "unknown" may mean one of a few things. LYQ may have screamed for help from anyone rather than a specific individual. Or he may have intended a specific individual, but because our observers lack the ability to read minds they could not identify

which individual was the intended recipient. "Unknown" does not mean that the observer simply doesn't know the name of the requested individual.

Let's talk about how these data should be entered. First, the column headings.

Column Name	Description
mid	A number that identifies the row within the uploaded file. The number must increase with each row and must be sequential.
coal_id	This number is an identifier which associates all lines within a single interaction. All rows associated with a given MPI must share the same number.
grp	Contains the group number where the animal was on the day of the interaction. See below for more specific information.
date	As above, use the format dd/mm/yyyy
actor	This column can contain more than one actor in instances where the interaction is not distinctly dyadic. If more than one actor, use a comma—and no spaces—to separate them.
agg_act	This is the "act". See below for more information.
recip	This is the "actee". It can contain more than one actee in instances where the interaction is not distinctly dyadic. If more than one actee, use a comma—and no spaces—to separate them. See below for more information.

outcome	Contains responses for those rows with an act of “?”, and should be empty otherwise. See below for more specific information.
form_passive_aid	Leave blank. This column doesn't have a purpose, yet.
context	Indicates the context in which the interaction took place, if the observers noted anything unusual. This value should only be provided on the first line of the MPI; it must be blank on all other lines. When context_type (below) is “C”, this column must be “CONSORT”.
consort	Indicates the outcome of the MPI when the “context” is “CONSORT”. See below for more-specific information.
context_type	Contains a code for the “type” of context in which this multiparty interaction took place. Each MPI must have a context_type on its first line. When a consort is involved, the context_type is “C”. Otherwise, this should be “N”.

You may notice the absence of an “observer” column. Perhaps—hopefully—one day we’ll start recording the observer, but for now, we don’t.

The “Grp” column

Our database uses unique numbers for each group. So instead of the group’s name, you need to provide the group’s number. Below is a list of all the current study groups’ numbers.

Acacia’s: 1.211	Europa’s: 1.2222
Hokey’s: 1.11	Narasha’s: 2.11
	Yoda’s: 1.2221

Trailing zeroes are okay to add, but also not necessary. For example, 1.11, 1.110, and 1.1100 are all acceptable ways to refer to Hokey’s group.

The “Agg_Act” column

MPI’s may use any of the acts used for agonisms and groomings. There are several other agonism codes that are exclusive to MPI’s that you may see. These are too numerous to include here, but they’re fairly similar to the ones you already know. If you see a code you don’t recognize, enter it just as the observer entered it, or ask the data manager.

There also a few new acts that are not agonism-related and occur frequently in MPI’s. These acts are:

- ? Actor asked actee for help
- + Actor helped actee
- P Actor gives **P**assive help to actee

The “Recip” column

In lines where help is requested (lines with act “?”), the actee is often unknown, or “UNK”. In these cases, enter “U”, not “UNK”.

The “Outcome” column

In the “?” rows, the observer may write a short code after the actee, indicating how the actee responded to the request for help. When recording these “outcomes” of help requests, you will enter something slightly different from what the observer actually wrote. See below.

Outcomes

Code	Meaning	How it is recorded
+	Actee provided help	SUCC
-	Actee did not provide help	FAIL
P	Actee provided passive help	PASS

When the actee is “UNK”—which should only ever happen in a “?” row—the “outcome” must be entered as “FAIL”.

The “Consort” column

When the observer notes that the MPI was part of a consort, s/he will include comments indicating who was consorting whom, and whether he kept the consort or if another male took over the consort. Those data are recorded in this column using a very specific format. Any data in this column should always be of the format:

ADM WITH ADF;ADM GET ADF
or
ADM WITH ADF;ADM KEEP ADF
(No space after the semicolon!)

Any deviation from this structure, such as spaces, or different wording, will create an error.

Examples

See the below MPI sheet. It’s a bit old, from a time when MPIs were recorded on their own sheets instead of sharing a page with Mounts & Consorts, but otherwise nothing about the data has changed.

113 JKW

~~ACACIAS~~ MULTIPARTY INTERACTIONS

07 MAR 18	
CAX	CONSORT UNA
CAX	AS ACC
YOI	P ACC
OBL	P ACC
(CAX KEPT CONSORT)	
30 MAR 18	
EID	AS IGL
ERM	+ EID
ACA	AS OPA
ACI	+ ACA
APP	+ ACA
AKP	+ ACA
ERM	+ ACA
EID	+ ACA
OBL	+ OPA

These data would be entered the data as shown below.

mid	coal_id	grp	date	actor	agg_act	recip	outcome	form_passive_aid	context	consort	context_type
1	1	1.211	07/03/2018	CAX	AS	ACC			CONSORT	CAX WITH UNA;CAX KEEP UNA	C
2	1	1.211	07/03/2018	YOI	P	ACC					
3	1	1.211	07/03/2018	OBL	P	ACC					
4	2	1.211	30/03/2018	EID	AS	IGL					N
5	2	1.211	30/03/2018	ERM	+	EID					
6	3	1.211	30/03/2018	ACA	AS	OPA					N
7	3	1.211	30/03/2018	ACI	+	ACA					
8	3	1.211	30/03/2018	APP	+	ACA					
9	3	1.211	30/03/2018	AKP	+	ACA					
10	3	1.211	30/03/2018	ERM	+	ACA					
11	3	1.211	30/03/2018	EID	+	ACA					
12	3	1.211	30/03/2018	OBL	+	OPA					

Validation and Update

After the separate copies have been compared and the "B" copy is deleted, the remaining compared and corrected data file is the "A" copy. When uploading the data to Babase, the system will validate the data by checking it against the

database's numerous rules, e.g. individuals that are not in the demography data or that are recorded as deceased as of that date.

1. Test and validate your data in a "test" database. Don't upload it to the actual database until you've finished validating the data. See the [Babase Wiki](#) for info about copying to a test database.
2. Go to the list of [Program Links](#) on the Babase wiki and open the necessary upload program:
 - a. For agonisms/groomings and mounts/consorts, use the generalized upload program: "upload".
 - b. For multiparty interactions, use the specialized MPI uploader: "upmpi".
3. Choose the database you're loading data into. If it's a test database (e.g. "babase_test"), then you'd provide that database's name. When ready to add to the actual production version of Babase, the database is simply "babase".
4. Login using your Babase username and password.
5. If uploading ad-libs or mounts/consorts and using the generalized uploader, the "Table or View" you'll load into is ACTOR_ACTEES. When uploading multiparty interactions, this information is not requested.
6. For ad-libs and mounts/consorts, check the box that allows the upload of NULL values and leave the NULL representation blank. This tells the uploader that any blank cells should be interpreted as "NULL" values. Again, this is not part of the multiparty interactions uploader.
7. Use the "Browse" button to select the tab-delimited text file you wish to validate.
8. Run the program by clicking "Upload". This will not actually upload the file if there are any errors.
9. When there are no errors, the update will be successful and the program will tell you how many records it added.

Be aware that some validation errors may require a bit of research before they can be resolved. For example, an individual who is recorded as deceased according to the demography data may turn up in a grooming. This may be because the individual never died, the three letter name code may have been recorded incorrectly on the original, or the animal may have been misidentified in the field. In any case, the discrepancy must be resolved. Ask Susan, Jeanne, or email the observers. With time, you will acquire a sense to know for yourself how to resolve the issue, or if you should just delete it.

Keep running the program until no errors are left. After all the issues have been resolved, the dataset is ready to be added to the database. **Remember that you CANNOT update the interaction data until the demography update has been completed at Princeton through the time period that the interactions were recorded. Interaction data gets uploaded quarterly, after the demography update has been done.**

10. Copy the “A” file to the appropriate “Final Data” folder
11. Make a note in the logbook as you go

Checking babase for interaction errors

Once the update for a given time period (say, quarterly) have been made, use the ACTOR_ACTEES to check for errors. Common errors include cases where the actor and actee were in different groups at the time of the interaction. Use a query like the one below to identify errors.

(This query returns errors in data collected on or after 1 Oct 2011)

```
select *
  from actor_actees
 where date >= '2011-10-01'
       and actor_grp <> actee_grp
 order by date;
```

Comparing interaction data with census records

It would be ideal to include a rule requiring that interactions can only occur on dates where the group was actually censused. This is not implemented as a rule for a variety of reasons, including but not limited to 1) some interactions were originally recorded as being on the first of the month instead of the actual interaction date; 2) individuals are sometimes censused and interacting in multiple groups in a single day but can only be censused in Babase in one of them; and 3) graduate students collecting data in the field rarely record complete censuses, so their incomplete censuses may simply not be recorded in Babase at all. All of these issues can be avoided or overcome; some day in the future, we may deal with these and officially add this rule.

There are certainly some cases where this happens for reasons that can't be explained by any of the above reasons. They are almost certainly errors, and will take some judgment to determine if they can be corrected or if they should just be deleted. An example way to query for these is:

```
SELECT *
  FROM actor_actees aa
 WHERE act = 'G'
       AND date > '2006-07-01'
       AND NOT EXISTS(select *
                      from census
                      where date = actor_actees.date
                            and grp in (actor_grp
                                         , actee_grp))
 ORDER BY iid;
```

This query shows all groomings since 1 July 2006 that were recorded on a date where neither actor_grp nor actee_grp was censused.

Historical note

In Feb 2017, it came to our attention that this was an issue for a set of ~1200 groomings and ~1500 agonisms since July 2006. On close inspection, we determined that many (not all) of these are because they were collected by Courtney Fitzpatrick (“CLF”), a graduate student at the time whose interactions were added to Babase but not her incomplete censuses. The number of interactions that don’t fit into this set is hard to say but is certainly small compared to all the data from this period. We decided in a Babase conference call that the time and effort required to backfill CLF’s censuses and then go sleuth through the other interactions were not worth the presumably-small gain in data cleanliness that we’d receive. So we’re leaving this alone for now. Going forward, the Duke data manager should watch for these issues in each quarterly update and keep the problem from continuing to occur, but correcting them in older quarters is not a priority right now.

FOCAL SAMPLING DATA

The focal sampling data is also referred to as the point samples and the focal animal samples data. These data have been collected since the project began but were handwritten until August 1999. From then through July 2015, the data were collected with handheld Psion Workabout units. The first two months are riddled with errors, so those data were discarded. Therefore, Psion data effectively began in October 1999. Beginning in ~May 2015, the team began testing the Android app AmboPrim8, a customized version of Prim8 (<http://www.prim8software.com/>). By the end of July, they had become proficient enough with their new Samsung tablets that we felt their data were reliable, so Psion data ends in July 2015 and “Samsung data” begins August 1, 2015.

Focal data is emailed from the field every day, or sometimes every other day. (See section on filing Amboseli data for more information). AmboPrim8 exports its data into 9 files, only one of which do we really need but we have the team send them all just in case. They’re saved in a folder that should be named with the date, the observer initials, and the tablet that collected the data, using the format “YYYYMMDD OBSERVER TABLET”. So JKW’s data on tablet A from 8 Jun 2018 is all in a folder called “20180608 JKW SA”.

When you receive these data from the field, you should save one copy of the folder in *[Your Dropbox]/Alberts Lab/ABRP_Data Management/Data from Amboseli/[Directories sorted by date]/Focal Sampling*. You should also save a “working” version (the version that you’ll work on, check, edit, etc.) in *[Your Dropbox]/Alberts Lab/ABRP_Data Management/DATA/FOCAL SAMPLES/In Progress/[folders separated by month]*.

Proofing

The raw data from AmboPrim8 is rather...arcane. A program has been developed to 1) convert the data into a format similar to that used by the Psion software, 2) automate a lot of the common errors described below, and 3) write SQL to upload the data to Babase. For more information, see *[Your Dropbox]/Baboon Project General Shared/Prim8/Reading and Interpreting Prim8 Data.docx*. (This requires being added to that “Baboon Project General Shared” folder, which you should have access to as the Duke Data Manager.)

The program discussed in that document is still kind of a work in progress; when it is finalized (and probably hosted on Papio), the text of that document will likely be split between this protocol and Babase’s tech specs.

Before uploading the data to Babase, check each one and make sure the data looks okay. The Prim8 Interpreter (discussed in the file mentioned above) can check for common errors/issues and alert you to them. See “Reading and

Interpreting Prim8 Data” for a lengthy discussion of how to interpret (and fix, if needed) the alerts/errors that it flags.

Validation and Update

Focal data needs to be updated in coordination with the demography update. Once all files for a given quarter have been proofed (*and assuming the demography update has been completed at Princeton*), you can then update the information into Babase.

1. Before attempting to update anything, you should copy Babase into a test database, probably babase_test. You can attempt uploads and work through errors there, then re-upload error-free data to real Babase. See “Copying the Babase Schema” on [this page](#) of the Babase wiki.
2. Follow the instructions in “Reading and Interpreting Prim8 Data” to see where to go to attempt to upload focal data.
3. When any errors are identified and corrected, be sure to log these changes. See old logs in *[Your Dropbox]/Alberts Lab/ABRP_Data Management/DATA/FOCAL SAMPLES/Data Change Logs* for examples of how to do this.
4. When a month is uploaded, the folder should be moved to the ‘Final Data’ folder.

Common Errors

When focal data are uploaded into Babase, there are some errors that are especially common. These are mostly addressed in “Reading and Interpreting Prim8 Data”. Below is a list of rules for handling such errors, decided by SCA and JA in October 2006. Many of these are errors that occur in Psion data but not Prim8, but some apply to both. Also, these rules can provide a helpful sense of how you should approach correcting errors in these data in case you find an error that is not addressed below.

1. When the error message states that an individual is its **own neighbor**, the neighbor should be changed to 998.
2. When the error message states that there is a **‘non-unique’ neighbor**, the second occurrence of the neighbor on that line should be changed to 998.
3. When the error message states that there is a **dead neighbor**, biograph should be checked to ensure that the neighbor is dead. It is also helpful to make checks to ensure another animal wasn’t the intended neighbor, although this judgment cannot usually be made. In most cases, the neighbor should be changed to 998.
4. When the error message states **‘neighbor not found in biograph’**, it can be useful to check the file, to ensure that a typo wasn’t made. There are

- cases in which the intended neighbor is very clear, and this can be fixed. In other cases, the neighbor should be changed to 998.
5. When the error message is **invalid sname in samples** or in other words, the focal animal doesn't exist in biograph, the sample should be deleted unless the intended animal can be discerned. This can be sorted out by checking the teams hand written focal sample logs which list each animal sampled on a given day.
 6. When the error message is **jpsamps after first conception** the sample should be deleted. This usually means that she was sampled as a juvenile and not an adult, and different types of information are collected for the two.
 7. When the error is **Invalid time**, the offending line(s) should be deleted as the times are limited to after 07:00 and before 19:00.
 8. When the error is **missing foodcode** the foodcode should be entered as UNK.
 9. When the error is **invalid foodcode** the foodcode should be entered as UNK.
 10. When the error is **male in fpoints** the points should be deleted.
 11. When the error is **Invalid Posture**, the line with the invalid posture should be deleted.
 12. When the error is **Foodcode Present but no 'F' in activity**, the foodcode should be removed.
 13. When the error is **invalid neighbor code** (for example 988 or 999) the neighbor code should be changed to 998.
 14. When the error is **Juv sample after matured** the sample should be deleted, unless it can be discerned that the sample should have been for a different individual.
 15. When the error is **Juv sampled as Female** the sample should be deleted, unless it can be discerned that the sample should have been for a different individual.
 16. When the error is **The actor and the actee on PARTS cannot be the same individual (errors for adlibs collected in Psion)**. The same should be deleted.

GPS/SWERB DATA

Data on home range and group movement are collected through SWERB records and GPS readings. SWERB files date back to the early 1980's and collection continues today (see Monitoring Guide for a description of data collection). During this time period, however, the specific protocols for data collection and entry have changed somewhat. This is an important to consider when using older data files. Until January 2004 data were hand-written; in January 2004 the field team started using handheld GPS units that are capable of storing the GPS data in digital form and downloading it to a computer. The handwritten data were entered in Foxpro (DBF files); now, electronic GPS files are received as text (.txt) files and imported to Babase. Older files, between Jan 2004 and July 2006, were collected as Mapsource files (.mps). All of the handwritten SWERB data has been entered, so we only are concerned about new GPS data from the field. GPS data is emailed from the field every day, or sometimes every other day. (See section on filing Amboseli data for more information). These data files should be added to the database once or twice each week, if possible.

Processing data from the field

Step 1: Save files from Email

Each observer emails their data every day, or every other day. The GPS data in these emails—like all the other data they send—should be downloaded and archived immediately. These original GPS files are saved, by month, in the appropriate *ABRP_Data Management/Data from Amboseli/[Directories sorted by date]/GPS* directory. Edited copies are saved in the appropriate year/month folder in *ABRP_Data Management /DATA/GPS/In Progress*.

Step 2: Remove useless information

The .txt GPS files include a lot of text automatically generated by the GPS handheld unit. This text is useless to us and should be removed. When you open the .txt file in Excel, it will look something like this:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Grid	UTM										
2	Datum	WGS 84										
3												
4	Header	Name	Description	Type	Position	Altitude	Depth	Proximity	Display Mod	Color	Symbol	Facility
5												
6	Waypoint	D1400	#####	User Waypoi	37 M 289476	1139.5 m			Symbol & Na	Unknown	Waypoint	
7	Waypoint	D1430	#####	User Waypoi	37 M 289400	1143.6 m			Symbol & Na	Unknown	Waypoint	
8	Waypoint	D1500	#####	User Waypoi	37 M 289340	1154.4 m			Symbol & Na	Unknown	Waypoint	
9	Waypoint	D1530	#####	User Waypoi	37 M 289305	1151.5 m			Symbol & Na	Unknown	Waypoint	
10	Waypoint	D1730	#####	User Waypoi	37 M 288607	1152.7 m			Symbol & Na	Unknown	Waypoint	
11	Waypoint	D1800	#####	User Waypoi	37 M 287968	1143.4 m			Symbol & Na	Unknown	Waypoint	
12	Waypoint	DBUNK	#####	User Waypoi	37 M 289317	1139.8 m			Symbol & Na	Unknown	Waypoint	
13	Waypoint	DDEILS	#####	User Waypoi	37 M 301770	1141.7 m			Symbol & Na	Unknown	Waypoint	
14	Waypoint	DEJKW	#####	User Waypoi	37 M 301769	1142.2 m			Symbol & Na	Unknown	Waypoint	
15	Waypoint	DEUNK	#####	User Waypoi	37 M 289332	1151.5 m			Symbol & Na	Unknown	Waypoint	
16	Waypoint	DMATAD	#####	User Waypoi	37 M 289235	1153.2 m			Symbol & Na	Unknown	Waypoint	
17	Waypoint	DMDTBA	#####	User Waypoi	37 M 289316	1145.0 m			Symbol & Na	Unknown	Waypoint	
18	Waypoint	G1630	#####	User Waypoi	37 M 289423	1150.1 m			Symbol & Na	Unknown	Waypoint	
19	Waypoint	G1700	#####	User Waypoi	37 M 289360	1150.1 m			Symbol & Na	Unknown	Waypoint	
20	Waypoint	GBUNK	#####	User Waypoi	37 M 289337	1152.0 m			Symbol & Na	Unknown	Waypoint	
21	Waypoint	GEUNK	#####	User Waypoi	37 M 287792	1142.2 m			Symbol & Na	Unknown	Waypoint	
22	Waypoint	GMATAD	#####	User Waypoi	37 M 287699	1143.6 m			Symbol & Na	Unknown	Waypoint	
23	Waypoint	GMDTBA	#####	User Waypoi	37 M 289337	1149.1 m			Symbol & Na	Unknown	Waypoint	
24												

To remove the useless text, delete rows 1, 2, 3 and 5. Next, delete columns A, D, and G-O (keep only columns B, C, E-F).

Step 3: General prep for the database

To put the data in a form recognizable by the database, change the remaining column headings to all-lowercase. Next, sort the data by Column B ("description"). Change the format of the "description" column to "yyyy/mm/dd hh:mm".

Add an "is_effort" column to the end. For all "begin" rows, their "is_effort" should be TRUE. All other rows should keep this column empty.

Creating a macro or script to clean up these files can streamline this process considerably.

	A	B	C	D	E
1	name	description	position	altitude	is_effort
2	DDEILS	2014/05/01 12:29	37 M 301770 9703051	1141.7 m	
3	DEJKW	2014/05/01 12:30	37 M 301769 9703053	1142.2 m	
4	DBUNK	2014/05/01 13:50	37 M 289317 9697371	1139.8 m	TRUE
5	DMDTBA	2014/05/01 13:51	37 M 289316 9697371	1145.0 m	
6	D1400	2014/05/01 14:00	37 M 289476 9697170	1139.5 m	
7	D1430	2014/05/01 14:30	37 M 289400 9696858	1143.6 m	
8	D1500	2014/05/01 15:02	37 M 289340 9696922	1154.4 m	
9	D1530	2014/05/01 15:33	37 M 289305 9696979	1151.5 m	
10	DEUNK	2014/05/01 16:04	37 M 289232 9697035	1151.5 m	
11	DMATAD	2014/05/01 16:04	37 M 289235 9697037	1153.2 m	
12	GBUNK	2014/05/01 16:07	37 M 289337 9696778	1152.0 m	TRUE
13	GMDTBA	2014/05/01 16:08	37 M 289337 9696778	1149.1 m	
14	G1630	2014/05/01 16:30	37 M 289423 9696747	1150.1 m	
15	G1700	2014/05/01 17:00	37 M 289360 9697081	1150.1 m	
16	D1730	2014/05/01 17:30	37 M 288607 9698663	1152.7 m	
17	D1800	2014/05/01 18:02	37 M 287968 9698937	1143.4 m	
18	GEUNK	2014/05/01 18:34	37 M 287792 9699016	1142.2 m	
19	GMATAD	2014/05/01 18:35	37 M 287699 9699079	1143.6 m	
20					

A final GPS file, ready for more-specific proofing and correcting, is shown above.

Step 4: Proofing and Correcting

It will be tremendously helpful to the Duke Data Manager if they familiarize themselves with the SWERB/GPS section in the Monitoring Guide, as much of the correction requires understanding what these data codes mean and how they're collected.

Data should be checked for errors regularly so we can get back to the field team about them as soon as possible. The team is much more likely to be able to solve inconsistencies if we get back to them right away. Record proofing, errors, and corrections in the GPS logbook.

GPS files may contain the following entries:

- Departures for both Drivers and Observers (should always be present)
- Begin (will almost always be present)
- Median Descent Time, or MDT (required, if Begin is present)
- Half Hourlies (will almost always be present)
- End (will almost always be present)
- Median Ascent Time, or MAT (required, if End is present)
- Lone Animal Observations (rarely present)
- Other Groups Observations (commonly present)
- Water Hole Observations (commonly present)
- Predator Sightings (commonly present)

See the Monitoring Guide for an explanation of conventions for recording these entries.

General guidelines for proofing GPS data:

To get the file in a format that the database can understand, there are some modifications you may need to make to the data. The below modifications don't actually change the data in any way, so they don't need to be documented in a change log.

- Departure rows need to be given observer-first, driver-second. Most of the time, the driver's row occurs first chronologically, so after sorting by "description" above, these rows will probably be out of order.
- MDT rows need to come immediately after Begin rows (and MAT's immediately after Ends). If chronological sorting puts some data between the two entries, cut and paste the MDT/MAT to just after the Begin/End.
- If the team notes a lone individual (usually a lone male)—indicated by beginning a line with "LM"—add a column called "lone_animal". For the lone

individual line, the correct value for this column is a single digit indicating the individual's sex: M or F. Leave the column empty for all other lines.

If you need to make any changes based on the below, these changes should be documented in a change log. In addition, many of these changes will require you to communicate with the team to know how to correct the error.

- Scan the date to make sure it is correct (same as the filename) and that only one date exists in the file.
- Check the Group.
 - Sometimes more than one group is monitored in one file, but not always. If two groups are monitored they should both have B and E records and MDT and MAT records. Pay close attention when multiple groups are recorded, as there are often errors involved with indicating the correct group. Clarify with the team if there is confusion.
- Check B and E records (Make sure groves are valid).
 - Grove ids are listed when the team knows or is pretty sure where the group slept. If they do not know or it is a new grove the entries will be UNK or NG respectively. Leave the NG entries as-is, and wait for follow-up from the team about the new grove's name before importing the file. The team will start to use the new name once it has been decided upon.
 - If a name for the new grove is not decided, e.g. the grove is not ever used again, you can replace the “NG” with “UNK” to upload it.
 - In case of multiple B's and E's for the same group, the team will likely include numerals after the B/E but before the grove (if any). Remove these unnecessary numerals.
- Check MDT and MAT records
 - For each group in one file, there can only be one MDT and one MAT. If the team provided more than one (for example, they ended observation to watch a different group, then came back later), the middle ones must be fictional MDT's/MAT's, and must therefore be removed. Retain the begins and ends, but remove these middle-of-observation MDTs and MAT's. Keep the MDT at the very beginning of the group's observation for that day, and the MAT at the very end of observation for that day.
 - One exception to this rule is when subgroups sleep in separate groves, and the team manages to record the MDT/MAT for both. This requires some special treatment. (See "Correcting Errors and Special Cases in GPS Data", below)

- Check ½ hourly records
 - Make sure the name column indicates an actual half hour, and is within a few minutes of the timestamp recorded by the GPS unit. Make sure there are no gaps. Missing half hourly records cannot usually be recreated but are noted as errors and mentioned to the team.
- Check W records
 - Similar to grove ID's, make sure there is either a RAIN or waterhole entry. Make sure that the waterhole name is valid. If it is a new waterhole the id will be NW. We leave these as NW and do not import the data file until the team agrees upon a name
- Check OG rows
 - These indicate a sighting of an “Other Group”. This may be non-study groups that we've named, or it may be a different study group from the one currently under observation. After the “OG” should be 3 characters indicating the group seen (e.g. “ACA” for Acacia's group). Make sure these 3 characters are a valid value from the GROUPS.Three_Letter_Code column in Babase.
 - After the group's 3-letter abbreviation, there may be a number, indicating approximately how many meters away the group is from the GPS point. This number should be a multiple of 5. If the group is less than 50 m away, a number likely will not be recorded at all.

Common errors to watch for:

- Grove and waterhole id's that are not on our list
- Grove and waterhole names that are misspelled or wrong
- Descent and ascent times (MDT and MAT) that are mixed up or missing
- BA (in MDTBA) and AD (in MATAD) are mixed up, e.g. “MATBA” or “MDTAD”
- No departure lines
- Changes in which group is designated (i.e., wrong group typed in for some of the records)
- Missing ½ hourly records
- No begin or end records
- Forgotten W in a water record
- O's instead of 0's (and vice versa), 1's instead of l's (and vice versa)
- Errors in B1 and E1 entries
- Altitude recorded in feet instead of meters.

Record all errors and corrections in the change log. Most errors will be very obvious and can be corrected without assistance. For those that aren't so obvious, send an email to the field team by the end of the day on Friday. They usually check and respond to emails on Saturday or Sunday. They will respond with corrections.

Correcting Errors and Special Cases in GPS Data

Missing B or E record

If a B or E record is missing and the team cannot give any details then add a row with BMISS or EMISS after the group letter code. Use the coordinates and time from the MDT record if the B record is missing or from the MAT record if the E record is missing. If both MDT and B (or MAT and E) records are missing, enter BMISS and MDTDU (or EMISS and MATAU), and leave all other columns empty. Add two new columns: "date" and "source". For both the BMISS and MDTDU, enter the date in the date column. For the BMISS (or EMISS) but not MDTDU (or MATAU), enter NR ("not possible to recover") as the "source". Leave "date" and "source" empty for all other lines in the file.

Missing MDT or MAT record

If an MDT or MAT record is missing and the team cannot give any details then add a row with MDTMISS or MATMISS after the group letter code. Use the coordinates and time from the B record if the MDT record is missing or the E record if the MAT record is missing. If both MDT and B (or MAT and E) records are missing, see the description above in "Missing B or E record".

Incorrect grove

When a grove is an incorrect code and the team cannot give details then we replace the grove name with UNK. The rest of the entry is assumed correct. For example, if the B entry was listed as VB98 and the team does not know if this is grove 98N or 98S then we would change the entry to VBUNK.

Missing or Incorrect Water Hole

When a water hole is an incorrect code and the team cannot give details then we replace the water hole name with MISS. The rest of the entry is assumed correct. For example, if the W entry was listed as VWCWE and the team does not know if this is CWE1 or CWE2 or CWEL then we would change the entry to VWMISS.

Missing departure record

Enter the missing Departure records as DMISS and DDMISS with all other columns empty. Add a "date" column and type in the appropriate date.

Subgroups in OG rows

Occasionally an OG row is indicated as a sighting of a subgroup of the indicated group, e.g. "OGVOGS75" is how the team would indicate that a subgroup of Vogue's group is ~75 meters away. This notation is not acceptable, because after reading "VOG", Babase will interpret the point as being "S75" meters away. To indicate a subgroup in an OG row, move the S between the O and G, e.g. "OSGVOG75".

The ability to import "other" subgroups this way has been in place as long as the SWERB tables have been up and running in Babase, but for years this fact was unknown to data management. In the data between ??? and ~January 2018, the subgroup "S" was removed from the file before upload. It was presumed that this was a minor detail, and not a significant loss. At some point in the future, we may decide to attempt to backfill these rare cases (they usually only occur in study groups, just before a group fission).

Unknown Lone Animal

If the team indicates a lone unknown animal ("LMUNK"), treat this entry like any other lone animal, as indicated above, with two small changes. Replace the sname "UNK" with "998", and in the "lone_animal" column, indicate the sex as 'M'. These are both a little odd, but they fit with other systems in place in babase that handle the recording of an unknown animal, so we do this to adhere to the convention.

If an unknown female is sighted (rare but not unheard of), the team will probably note it with "LF" instead of "LM". You should change it back to "LM", and indicate "F" (the individual's sex) in the "lone_animal" column that you added.

Multiple MDT's/MAT's with Subgroups

Very rarely, the team might record separate descents or ascents for separate subgroups. For example, one subgroup might have an MDT from grove A at 6:30, while the other subgroup has an MDT from grove B at 7:00. Recording this is a little weird.

Basically, we're going to treat one of the subgroups as though it were a completely different group by beginning and ending a very brief bout of observation, and we'll add a column that the database will need to make this work. Let's say your data records the following:

name	description	position	altitude
------	-------------	----------	----------

VSB151	2012/06/07 6:35	37 M 287847 9699731	1156.6
VSMDT0649	2012/06/07 6:49	37 M 288566 9699667	1139
VSB98N	2012/06/07 6:39	37 M 288564 9699668	1142.6

The team failed to provide an MDT for the second subgroup, but we do know the grove that the subgroup descended from, so we'll record it. First, let's add an extra MDT row, following the rules for missed MDT's described above:

name	description	position	altitude
VSB151	2012/06/07 6:35	37 M 287847 9699731	1156.6
VSMDT0649	2012/06/07 6:49	37 M 288566 9699667	1139
VSB98N	2012/06/07 6:39	37 M 288564 9699668	1142.6
VSMDTMISS	2012/06/07 6:39	37 M 288564 9699668	1142.6

Next add a "secondary_ad" column to the data file. For the second (chronological) MDT/MAT, enter "TRUE" in both the begin/end and MDT/MAT lines.

name	description	position	altitude	secondary_ad
VSB151	2012/06/07 6:35	37 M 287847 9699731	1156.6	
VSMDT0649	2012/06/07 6:49	37 M 288566 9699667	1139	
VSB98N	2012/06/07 6:39	37 M 288564 9699668	1142.6	TRUE
VSMDTMISS	2012/06/07 6:39	37 M 288564 9699668	1142.6	TRUE

Finally, add a fictional end line and MAT, setting the date/time and GPS coordinates to the same as the begin line, and set "secondary_ad" to "TRUE", as well:

name	description	position	altitude	secondary_ad
VSB151	2012/06/07 6:35	37 M 287847 9699731	1156.6	
VSMDT0649	2012/06/07 6:49	37 M 288566 9699667	1139	
VSB98N	2012/06/07 6:39	37 M 288564 9699668	1142.6	TRUE
VSMDTMISS	2012/06/07 6:39	37 M 288564 9699668	1142.6	TRUE
VSE	2012/06/07 6:39	37 M 288564 9699668	1142.6	TRUE
VSMATMISS	2012/06/07 6:39	37 M 288564 9699668	1142.6	TRUE

Groves and Waterholes

The master copy of the grove and waterhole IDs are kept in Babase in the SWERB_GWS table. When NG's or NW's show up in the data the team is supposed to write down the coordinates and name in the demography notes. When new monthly data comes, scan the demography notes and monthly report for new groves and waterholes and add them to the table. When a new grove or waterhole name starts showing up and you do not have the demography notes yet, you can email the team and ask if it is valid. Add it to the SWERB_GWS table and try to get a GPS reading from the team. If they provide you with coordinates for a grove, add them to the SWERB_GW_LOC_DATA table (upload using the SWERB_GW_LOC_DATA_XY view).

Other SWERB/GPS rules

The SWERB data are almost certainly the most complicated set of data that the Duke manager oversees. There are A LOT of other rules not addressed here, governing how the data should be entered, how things can and can't be recorded, etc. There is just too much for it all to be recorded here. Consider looking at recent years' GPS change logs to see how things were changed in the past, if you need help. The tech specs in Babase can also be helpful, but often they're too arcane to be comprehensible. If you struggle, hang in there! With time and familiarity, something will "click" for you and it will all make a lot more sense.

Data Management in Babase

Data Upload

Upload edited .txt files to Babase using the [upload program](#) listed in the [Babase Wiki](#). To import data correctly, all files should be uploaded to the SWERB_UPLOAD view. This view exists solely for the purpose of uploading SWERB data to the many SWERB-related tables.

Deleting/Updating Data

The SWERB data are recorded in so many interrelated tables that in most cases, corrections cannot be made via updating. To make changes, it's necessary to delete the old data, correct the data file, and re-import.

To delete SWERB data, delete from the SWERB_DEPARTS view. To make sure data are removed cleanly, it's best to delete by date. For example, if you enter:

```
delete from swerb_departs where date='2013-01-01';
```

All SWERB data recorded on 2013-01-01 will be removed. If you use this command to update a data file, make sure to re-import ALL of the files from the deleted date, not just the file you edited!

SWERB_LOC_DATA_CONFIDENCES

An R script has been written by David Jansen at Notre Dame to check groves (and someday, maybe water holes) indicated by the observers and compare the indicated coordinates with known coordinates for the groves. The script assigns a score indicating our "confidence" that the indicated grove really is correct, and these confidences are stored in SWERB_LOC_DATA_CONFIDENCES. We hope that someday in the future, this script can be rewritten in such a way that Babase will perform the check automatically when data are added to Babase, but

we have not yet attained this paradisiacal utopia. In the meantime, the Duke manager should periodically (quarterly, probably) run the script to add new confidence scores for not-yet-analyzed data.

The script is available at https://github.com/ArchieLab/checking_swerb_data, but be aware that this repository is private. An admin in the Archie Lab will need to grant you permission to see it. There's a README file there with info about how to run it. When finished, the script will output a new set of data for you to import to SWERB_LOC_DATA_CONFIDENCES. At the time of this writing, the script's output file includes some extra columns that need to be removed before upload, but besides that there's nothing else that needs to be done before uploading the data.

DARTING DATA

Data Storage

Darting data will come in each summer (~May-Aug), after that year's "darting season". You may also receive one or two darting sheets outside of that period, but these are rare. These data should be entered as soon as they come in with the scanned-and-emailed monthly data and filed in the "**Darting Photocopies**" binder. The original sheets will arrive with the rest of the original data and should be checked against the scans to ensure no notes were cut off during scanning. They should then be filed in the "**Darting Originals**" binder.

Data Entry

The data need to be entered with specific columns, which also need to be in a specific order. To ease this, templates are saved in **ABRP_Data Management /DATA/Darting/Darting Templates**. In case there's any issues with these, consult the documentation for our darting data uploaders at https://papio.biology.duke.edu/babase_system.html#Updart.

At the time of this writing (June 2018), there is no specialized uploader for the Dart Samples data. There once was, but in March 2017 we updated the tables used to store these data such that the old uploader would no longer work. Creating a new upload program for the samples data—or an uploading view—is a longstanding to-do item that we will hopefully, someday, maybe, probably, actually do. But for now those data have to be added manually.

Step 1: Open the darting template for each of the 7 different types of data with Excel. Immediately 'Save-as' and name the file using the following convention:

Dart_datatype_YYMM.txt

where datatype = logistics
morpho
samples
physio
teeth
ticks
testes

And 'YY' is the 2 digit year, and 'MM' is a two digit month.

Step 2: As you enter each line of data, using a pencil, place a small checkmark next to that line of data on the scan. This will help you to ensure you do not forget any sections of data.

Step 3: Be sure to incorporate any notes on the darting sheet in their appropriate location. Any notes regarding collar removal or deployment should be placed in the comments section of dart logistics in the following format: GPS Collar Fitted: Frequency XXX.XXX. Notes indicating an event that did not allow testes measures to be taken should be placed in the comments section of dart_testes, etc.

Once the data have been entered by one person (ideally NOT the data manager), they should be saved in **Working Data/darting**. Another person (often, the data manager) should then proof these data manually by checking each record line by line against the scanned darting sheets.

Data upload

The darting data are uploaded using datatype-specific upload programs that can be found at <https://papiro.biology.duke.edu/babasewiki/ProgramLinks>. Each upload program corresponds to the data type in its name. There are two different upload programs for the dart_testes data. The historical darting data on testes were collected more as an arc, or roughly $\frac{1}{2}$ of the circumference measured. More recent measures, and all subsequent measures, are done using diameter instead. For all new data, use the program titled: Updart Darting Testes Diameter Data.

The upload program will return all errors found in the file with line numbers provided for each error. No data from the file will be uploaded until all errors are fixed.

PREDATION AND HUMAN DISTURBANCES

Until the end of 2017, these data were scanned and filed each month as the scans came in. The protocol for interpreting and entering handwritten data is included in earlier versions of the data manager's protocol, but not here.

Now, the field team enters these data in their tablets as individual surveys in the [QuickTapSurvey](#) app, and the data are sent to us whenever the team connects their tablets to the Internet and syncs their data (weekly, sometimes more often).

QuickTapSurvey Management

As the Duke Data Manager, you are the main person responsible for everything related to QuickTapSurvey. This primarily involves maintaining the surveys our observers use to record their data, and proper management/archival of the observers' responses to the surveys. You should have been given the administrator's name and password for the baboon project's account.

This protocol intentionally avoids going into very much detail about how to do things in the QuickTapSurvey web interface. These kinds of details are surely subject to change over time, so trying to keep up with it here seems like a futile exercise.

Survey Upkeep

The surveys used by the team are pretty well-established by now, so you shouldn't really need to change or update any surveys unless there has been a group fission/fusion. In this case, you'll need to remove the old group(s) and add the new one(s).

Once a survey has been "published"—meaning it is available to accept responses—it can no longer be edited. This is annoying, but that's how it works. So if you want to "edit" a survey, you really will have to make a new one and edit its contents before publishing it.

If you're just trying to make some quick edits, don't try to remake the surveys from scratch! When looking at your list of surveys, you should have the option to "Duplicate" an already-published survey. This will create a "draft" version of the survey, which you are free to edit however you like until you decide to "publish" it.

When publishing new surveys, be sure to give them new names that will clearly emphasize what each survey is (Predation By Baboons, Predator Sightings, or Human Disturbances), and also will clearly emphasize that this survey is the new one that the observers should use now. You may be able to coordinate this transition with the team really well, but likely there will be a period of overlap

where they have both “new” and “old” surveys available to them. Clear names will help make it clear to them which surveys they should be using.

After you publish a new survey, be sure to “assign” the survey to all of our users. If you don’t explicitly grant permission for an observer to see the survey, s/he won’t see it.

Data Management and Archival

The web interface for viewing the data on quicktapsurvey.com is a little clunky. Ideally you should only use the website to download the survey responses, then use Excel to actually view them.

To make viewing the data easier for everyone, and to keep an archive in case we ever stop using QuickTapSurvey, all of the survey responses are saved in a shared Dropbox folder (shared with project leaders and data managers only) called “Predation-Human Disturbances”. Data are sorted by year, with separate directories provided for storing each kind of data. To keep file sizes manageable, use a separate file for each month, and name them using the convention [YET TO BE DETERMINED AT THE TIME OF THIS WRITING].

To keep the data spreadsheets semi-current, you should update them regularly. More regularly than once a month; probably 1-2X/week. (Doing it this often will also help you pay attention to their data and stay on top of any issues that may come up.) QuickTapSurvey gives you the option to filter by date which surveys you want to download, so it shouldn’t be difficult to download only the data which have arisen since your last download.

Dates and Times from QuickTapSurvey

The content of each of the survey types will vary, but all data downloaded from QuickTapSurvey begins with the same two columns: “Date Collected” and “Date Sent”. The meaning of these columns should be pretty self-explanatory, but there are a few details worth noting here.

The Date Collected column shows the time and date that the survey was completed on the tablet. This date/time is taken from the tablet’s clock at the time of survey completion. This is important, because occasionally the date/time on the tablets gets messed up. If the team emails to alert you that the date/time on their focal samples is wrong (they’re moderately good about doing this), be aware that this will also affect this “Date Collected” column. With this in mind, the accuracy of the “Date Collected” is not very reliable.

The Date Sent column shows the time and date that the surveys were synced with QuickTapSurvey’s servers. This date/time are from QuickTapSurvey’s servers, not from the tablet’s clock. For this reason, the time of the “Date Sent” is

reasonably reliable. Unfortunately, it's not nearly as useful as knowing when the survey was actually collected.

Both columns are recorded in "YYYY-MM-DD [Eastern time] - nnnn" format. No matter where the survey was collected or submitted, it's always in Eastern time. There is no GOOD reason for this; it's just because QuickTapSurvey and its servers are based in Ontario, which is in the Eastern time zone. The "-nnnn" will be "-0400" during daylight saving time (~Apr-Nov), and "-0500" otherwise.

For example, if I'm in Kenya and complete a survey on 15 Jan 2018 at 17:12:34 (Kenya time, UTC+0300), this is 09:12:34 in Eastern (standard) time. The "Date Collected" for this survey will be "2018-01-15 09:12:34 - 0500". If I did it in July (during DST) at the same time (in Kenya), the "Date Collected" would be "2018-07-15 10:12:34 -0400".

Babase Data Management

The Predation and Human Disturbance Tables

The Predation and Human Disturbance data are stored in three separate tables, all of which are currently stored in the babase_pending schema. This means that these tables are not documented with the regular Babase tables and should therefore be documented here.

Predation_by (Predation by baboons)

This table records events where baboons are the predators, hence the predation is by the baboons. It includes the following columns:

Predation By	
p_byid	A unique identifier for each row of the table. This row is managed by the database and should be omitted during data entry.
predid	An ID number for each event. If a major event on a date has several smaller events spanning multiple rows, the same predID is assigned to all those rows. Usually, each new row of the table receives its own predID. This column is NOT managed by the database. When adding new rows to the table, you should lookup the highest predid in the table and number the new rows beginning with [that highest predid + 1]. Note that the lack of DB oversight of this column means that there is widespread reuse of predids in this table when the events are completely unrelated.
date	Date of occurrence. If no date is present, leave blank.
time	Time of occurrence. If no time is present, leave blank.
gid	Group ID (from GROUPS.GId) for the study group being observed.
spec_code	Code that identifies the species of the prey. Legal values discussed later.

spec_count	Number of individuals of the prey. Enter 99 if too many to count or unknown.
rec_code_N	Codes that partly indicate baboon behaviors but also how the human observer observed the predation (e.g. seen vs. only heard) N = numeral assigned to each column, starting from 1 through 4. Note: this column holds WAY too much information, and should be implemented differently when/if we ever consider moving the data into Babase.
quadrat	Map coordinates (usually a three letter/number combination). This info hasn't really been recorded since the advent of GPS technology.
loc_desc	Descriptive explanation of the location
notfail_fail	Whether or not predation was successful (i.e prey killed and/or eaten). 0=fail; 1=success.
fail_pred_sname_N	Snames of those baboon individuals involved in predation/chase but not necessarily successful in killing or obtaining meat (may be multiple individuals); N = numeral assigned to each sname starting at 1. New columns may be added as needed.
succpred_sname_N	Those baboon individuals involved in the actual killing and eating of meat (may be multiple individuals); N = numeral assigned to each sname starting at 1. New columns may be added as needed.
consumer_sname_N	Name(s) of any animal(s) who actually consumed meat; N = numeral assigned to each sname starting at 1. New columns may be added as needed.
comments	Any notes that provide more in-depth information about the event

Predation_on (Predator Sightings or Evidence)

Although the name suggests it, this table does not exclusively record events where the predation is on the baboons. More generally, this table records observed predator activity in general. All predations should be recorded here, except those where the predation was by baboons. A discussion of each of the columns:

Predation On	
p_onid	A unique identifier for each row of the table. This row is managed by the database and should be omitted during data entry.
predid	An ID number for each event. If a major event on a date has several smaller events spanning multiple rows, the same predID is assigned to all those rows. Usually, each new row of the table receives its own predID. This column is NOT managed by the database. When adding new rows to the table, you should lookup the highest predid in the table and number the new rows beginning with [that highest predid + 1]. Note that the lack of DB oversight of this column means that there is widespread reuse of predids in this table when the events are completely unrelated.
date	Date of occurrence. If no date is present, leave blank.
time	Time of occurrence. If no time is present, leave blank.

gid	Group ID (from GROUPS.GId) for the study group being observed.
spec_code	Code that identifies the species of the predator. Legal values discussed later.
spec_count	Number of predator individuals. Enter 99 if too many to count or unknown.
rec_code_N	Codes that partly indicate prey behaviors but moreso how the human observer observed the predation and what s/he actually observed at all. N = numeral assigned to each column, starting from 1 through 4. Note: this column holds WAY too much information, and should be implemented differently when/if we ever consider moving the data into Babase.
notfail_fail	Whether or not the disturbance led to mortality. 0=no prey mortality; 1=yes, prey mortality
quadrat	Map coordinates (usually a three letter/number combination). This info hasn't really been recorded since the advent of GPS technology.
loc_descr	Descriptive explanation of the location.
victim_sname_N	If baboon mortality occurs, the victim sname is recorded; N = numeral assigned to each sname starting at 1. New columns may be added as needed.
latitude	If GPS coordinates are provided, latitude goes here.
longitude	If GPS coordinates are provided, longitude goes here
altitude	If GPS coordinates are provided, altitude goes here.
accuracy	If GPS coordinates are provided, accuracy goes here.
comments	Any notes that provide more in-depth information about the event.

Human_disturb (Human Disturbances)

The human disturbance table provides information about human disturbance events that occur while we are watching the baboons. We do not attempt to record events that occurred while we were not there. This approach allows us to measure the rate at which human disturbance events occur. In this regard, it is somewhat different from the predation data, in which we do sometimes record evidence of predation events that occurred when we were not with the groups, and sometimes predator sightings when we are not with the baboons.

Human Disturbance	
p_hdid	A unique identifier for each row of the table. This row is managed by the database and should be omitted during data entry.
predid	An ID number for each event. If a major event on a date has several smaller events spanning multiple rows, the same predID is assigned to all those rows. Usually, each new row of the table receives its own predID. This column is NOT managed by the database. When adding new rows to the table, you should lookup the highest predid in the table and number the new rows beginning with [that highest predid + 1]. Note that the lack of DB oversight of this column means that there is widespread reuse of predids in this table when the events are completely unrelated.

date	Date of occurrence. If no date is present, leave blank.
time	Time of occurrence. If no time is present, leave blank.
gid	Group ID (from GROUPS.GId) for the study group being observed.
spec_code	Code that identifies the nature of the disturbance. Legal values discussed below.
spec_count	Number of individuals of [whatever is indicated by spec_code]. Enter 99 if too many to count, or unknown, or if the spec_code refers to something where a single number doesn't make sense.
rec_code_N	Codes that partly indicate baboon behaviors but also how the human observer observed the disturbance (e.g. seen vs. only heard). N = numeral assigned to each column, starting from 1 through 4. Note: this column holds WAY too much information, and should be implemented differently when/if we ever consider moving the data into Babase.
notfail_fail	Whether or not the disturbance led to baboon mortality. 0=no baboon mortality; 1=yes, baboon mortality
quadrat	Map coordinates (usually a three letter/number combination). This info hasn't really been recorded since the advent of GPS technology.
loc_descr	Descriptive explanation of the location
victim_sname	If baboon mortality occurs, the victim's sname
human_type	When the spec_code indicates that some kind of humans were involved, this indicates which kind (Maasai? Tourists? Observers?)
comments	Any notes that provide more in-depth information about the event

The legal values for all three tables' spec_code and rec_code columns are shown below. They are shown together and provided on a new page for ease of reference during data entry.

Entry Codes for Predation Tables

Predation by

Spec_code	Description
1	Reptile
2	Bird
3	Egg
4	Gazelle
5	Hare
6	Other
7	Vervet
98	Unknown

Rec_code	Description
1	Predation seen
2	Carcass eaten (carcass seen implicit)
3	Carcass seen
4	Heard but not seen
5	Blood on baboon fur
10	Scavenging
98	Other

Predation on

Spec_code	Description
1	Hyena
2	Lion(ess)
3	Snake
4	Jackal
5	Bird
6	Leopard
7	Cheetah
8	Serval
98	Other
99	Unknown

Rec_code	Description
1	Predation seen (predator seen implicit)
2	Predator Seen
3	Carcass Seen
4	Predator heard but not seen
5	Footprints
6	Baboon and Vervet Alarms
7	Baboon Only Alarms
8	Vervet Only Alarms
9	Baboons Ran/Fled
98	Other

Human disturb

Spec_code	Description
1	Humans
2	Maasai & Dogs*
3	Maasai & Livestock*
4	Dogs alone
5	Livestock alone
6	Baboons chased from garbage or consumed garbage*
7	Maasai & Dogs & Livestock*
98	Unknown

Rec_code	Description
1	Seen
2	Heard but not seen
3	Baboon alarm
4	Baboons ran or fled
98	Other
99	None of the above recorded by observers

*Use spec_count 99

Data Entry

The forms in QuickTapSurvey were intentionally written to fit pretty clearly with the columns of these tables. Also, the QuickTapSurvey forms themselves are occasionally subject to change. Because of these reasons, this protocol does not go into a detailed discussion of how to enter the data. It should be pretty self-explanatory. Just make sure your files have all the same column names (in lower case) as the table that they're intended for.

Save the data in files using the following naming system:

[Dataset]-MMYY.txt

Where "Dataset" could be HumanDist, Predation_By, or Predation_On.

Before uploading the data to Babase, these files should be saved in the **[Dropbox]/Alberts Lab/ABRP_Data Management/DATA/PREDATION/In Progress** directory. After proofing and uploading them, they should be saved in the **.../PREDATION/Final/[sorted by date]** directory.

Proofing Data

Because entry of these data is essentially some skillful copy/pasting, the process for proofing these data is not nearly as extensive as other data types. After pasting the data into files with the right columns and column names, you should wait a period of time (~a day, at least) and then come back and compare your file with the original survey responses. A line-by-line comparison isn't necessary; just do some spot checks to make sure you moved the data over correctly.

Upload Data

These data are uploaded using the generic upload program (found at <https://papiro.biology.duke.edu/babasewiki/ProgramLinks>) on the Babase wiki. When using this program, be sure to include the schema name (babase_pending) when specifying which table you're uploading to, e.g. "babase_pending.human_disturb", "babase_pending.predation_on", "babase_pending.predation_by".

Since there are currently no validation rules on the predation data, you should not encounter any data-specific errors. You may encounter 'Not Null' errors or data-type errors (numeric versus varchar) but these should be straightforward.

RANKS

Updating Ranks

Female ranks are calculated and entered at Princeton. Male ranks get updated at Duke. This is typically done with Susan. “Ranker”, a program designed for the project by a computer science student, is used for this process. We create the “All Male” ranks and they are handwritten in the **Male Agonism Binders**.

Ranker can be found and downloaded from the [Babase Wiki](#). Go to the bottom of the page and choose “Attachments”. There are many different “tgz” files listed here, including different versions of Ranker. Choose “get” next to the one you want (probably the most-recent one), and the download will begin. See that page on the wiki for more information about how to start the app.

How to use Ranker:

1. Once you have logged in, you are presented with a screen to create the ranking. Select the ranktype (usually ‘ALM’ for our uses), the group in question, and the start and end dates (YYYY-MM-DD) of the month you would like to rank. The start date = the first day of the month of interest, and the end date = the last day of the month of interest. If you use day 31 on a month that only has 30 days, the date will turn red to indicate you have made an invalid entry. It is important that the ranking start and end dates match the interaction start and end dates. Also, if you try to look at more than 1 month at a time, you will receive an error message. Ranker will not allow you to save rankings to babase that span more than 1 month. Finally, Click ‘Rank’.
2. Ranker will ask you if you would like to load the ranking that is already in babase if one exists, load the last available ranking before you dates if one doesn’t already exist, or load the ‘default’ ranking. The default ranking orders individuals in order of age, and this is not an ideal option to use in most cases. Always choose to load the ranking from babase if one already exists.
3. Once the matrix has loaded, you are able to drag and drop individuals along the horizontal axis of the matrix by simply clicking their sname and dragging them to their proper position. Simply hovering over an animal will display demographic information pertinent to ranking such as birth, matdate, dispersedate if applicable, and rankdate if applicable. Matgrp and bstatus are also displayed.
4. Once you have placed all individuals in their proper rank positions, select ‘File’ and ‘Save Ranking to Babase’. You will get a popup confirming that the ranking has saved. You can then close the ranking and move on to the next month.

Note: Ranks are typically decided or 'called' by Susan. She will then read them back to the data manager who writes them down on tabulation sheets in a handwritten matrix. Ensure when Susan is calling back the rankings to you that every month for every group is read from beginning to end to ensure that no immigrations, emigrations, births, or deaths are missed.

Decision Rules for Assigning Male Ranks

S. Alberts 11/94

Revised: 6/95

Last revised: 11/2011 by SCA

1. Start with the matrix for the last month for which ranks were assigned.
2. Look at the matrix for the first month for which ranks have not yet been assigned, ordered as for the previous month.
 - a. Check for new immigrant males, who will appear at the end of the list of individuals in the matrix as their default position.
 - b. Put new immigrant males below all adult males and above all subadult males.
 - c. If a male leaves in one month and then returns several months later, treat him as though he had never left, i.e., put him where he was the month he left.
 - d. When a male comes into a group and also leaves in the same month and has no interaction the procedure will be as follows:
 - i. rank according to age if NOT adult
 - ii. Place at the bottom of the list of adults IF adult
 - e. If a male immigrates and is very shy so that we don't see him interact for several months, we rank him in the first month we can and back date that rank to his immigration date. This does not apply in all cases, only in cases where he interacts with no one (or extremely rarely) for several months because of shyness, and then when we finally see him start interacting, he has a clearly defined rank position well above the bottom.
 - f. If a male is present for only a few days in a month (< 1 week), generally, place him at the bottom of the hierarchy unless he is interacting a lot and wins over many animals. If there are only a couple of interactions, keep him at the bottom if he stays less than a week. This is because, even if he wins over a few males, he was functionally not a relevant member of the group for most of the month.

3. Now, check all entries below the diagonal and rearrange the order so as to minimize these entries, with the following caveats.
 - a. For every “win” below the diagonal, check the corresponding “loss” box above the diagonal. If the values in each box are the same, do not record a change in rank. Otherwise, whichever box has the higher value determines the overall winner in the dyad, i.e., determines who is higher ranking.
 - b. Place new males over all males that they clearly win over in the first few months, even if there are multiple jumps in these months, and especially if the data are scanty. I assume that a new male will find his place in the hierarchy without necessarily working his way up systematically from the bottom, so I didn’t insist that he have wins over everyone I am going to place him over.
 - c. A 1-0 win with an adjacent upper neighbor in a single month is sufficient to indicate a rank reversal. Unless the upper neighbor won consistently for the 3-6 months prior to AND after the reversal. Adjacent males frequently flip back and forth from month to month; each flip is recorded as a rank reversal for the month.
 - d. A win where the cross-diagonal box is not 0 (i.e. a contested reversal in which, for instance, the lower ranking male won over the higher ranking male 4 times but lost to him 3 times) should not be counted as rank reversal unless it is confirmed in the following several months.
 - e. If a male wins only once over a male more than one rank above him (i.e. over more than his adjacent upper neighbor), don’t change his rank unless that win or a win over more males in between is confirmed in the next 1-3 months, or more if data scanty. One exception to this is when a male wins many times over a male above him, like 6 or more wins – in that case give him credit for the rise but watch both him and the male he won over carefully for the next few months to see if the loser is dropping precipitously or the winner is rising.

Another exception is when a maturing male wins over a male several ranks above him, especially if he then disperses before he interacts with anyone else. In this situation, give the male credit for the rank rise. In general, these young “rising” males get more credit for multiple “jumps” up the hierarchy than do older established males, especially if they are obviously on an upward trend anyway.

- f. Sometimes a male wins over a male more than one rank above him several times in a month, but over no males in between. In order for a lower ranking male to get credit for rising in this situation, either (1) he must win over a male in between in the next few months, (2) the higher ranking male must keep winning over the males in between during the same period that the lower ranking male is winning over him, or (3) the lower ranking male must be a maturing male as described above. Alternatively, for situations in which none of these things occur, and in which the higher ranking male loses to males in between in the following month, it may be that the higher ranking male is dropping in rank (see 4b below).
 - g. It is always helpful to look at the 6 months or so before and after a rank change to get an idea of where the male is going and whether he confirms the wins/losses he had in a particular month. However, confirmations are not always necessary to record a change, as indicated above.
4. There are some important differences between male rankings and female rankings
- a. Don't assume stability as strongly for males as for females – male ranks change more often. Yes, scanty data sometimes makes it frustrating, it is sometimes unnerving to watch a male rise over a couple of months on the basis of single wins over a series of upper neighbors, but I think reflects reality and so I do it.
 - b. Always think first in terms of males rising in rank, rather than falling. That is, for every rise there is a concomitant fall and vice versa – so focus on the rises. If a male is really going to fall in rank, as a result of an injury or something, you will usually see a whole slew of losses on the part of that male over a month or two. If he has a couple of losses in a single month, don't just score him as dropping in rank, instead watch the winners and see what they are up to.
 - c. Because the time course of wins and losses is important for males, 6- and 12- month summaries are not as useful for males as they are for females (although they can show general patterns). For instance, for females, If DUD wins 5 times over NYA and loses 4 times to NYA in the course of a year, DUD is considered to rank above NYA by the end of the year, and generally for the whole period. For males, in contrast, it depends totally on when the wins and losses occurred. This is because male ranks are so age-dependent – i.e. a male rises and then falls in his life, and we want to see that trajectory. Maturing males, for instance, get credit for

attaining adulthood when they first win over adults, and the time course of their rank rise is of interest.

5. Throughout the ranking process I look at what particular dyads are doing over the several months before and after the months I am ranking. This dyad-oriented checking is a good thing to do in addition to overall checking, because it makes you pay attention to how often particular dyads are being reinforced or challenged in their ordering.
6. I find it easiest to sit with a set of printed matrices, printed before you've done all of the ranking properly, and a sheet of columned paper. I write a new ranking for every month as well as the relevant agonisms and comings and goings for the month. Make final rankings in this way, enter them into the computer, and make final clean copies of everything.

Ranked “On” and “By” Dates

In the RANKDATES table in Babase, we record dates for when every male first attained rank, i.e. when he first ranked above an adult male. If he attained this rank in a study group, then we note it with a “Rstatus” of “O”, for “On”, as in “XXX first attained rank ON [date]”. Because of immigrations/emigrations and dropped study groups, we often don't observe a male's *first* rank attainment and we can only state with confidence that he had attained rank “by” a particular date.

For example: individual MMM is a new immigrant in a study group on 2012-06-20, and is estimated to be ~10 years old. Immediately upon entering the group, he is seen “winning” agonisms against other males in the group. Given his age, he certainly attained rank for the first time elsewhere, but because this is our first observation of him, we note that he was ranked BY 2015-06-01 (note that in RANKDATES and RANKS we always group interactions by month, so the 20 June entry is truncated to the first day of June).

Ranked “on” dates are assigned by us while doing the monthly ranks. Most ranked “by” dates are assigned by the Princeton data manager, but we do assign some of them, as discussed below. See Princeton's data management protocol for a more thorough discussion about when/how to assign ranked “BY” dates.

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

Note: The wording of these rules predates adding residency to the MEMBERS table, so anywhere it says resident, it just means “present per MEMBERS”.

1. If the male is natal to our study groups and the date on which he first attained adult rank is known, he receives an “ON” designation (we know he ranked ON that date).

2. If the male is an immigrant male not natal to our study groups:
 - 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, then when he attains adult rank he should be ranked ON that date.
 - 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), he should be ranked "BY" 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 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 and before rank attainment, but is away from our observations and in an unknown location for a long time (more than a few weeks). We assign the male a ranked "BY" date that equals the first day of the month in which he first appears in members in the non-natal study group.
5. If the male is a natal male from one of our study groups and attained one or both markers (rankdate and maturedate) 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 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 (rankdate and maturedate), and attains one or both markers in a non-study group. We follow these rules:
 - a. 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:
 - a. 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.

- b. 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.

Wiki and Other Online Management Tasks

The Data Managers at both Duke and Princeton are responsible for upkeep of various parts of the Babase Wiki (<http://papio.biology.duke.edu/babasewiki>).

Updating Dataset Status Table

You need to be either a "WikiAdmin" or "WikiEditor" to do this.

Whenever new sets of data are added, especially after quarterly updates, the [Dataset Status Table](#) should be updated to reflect this. Once logged in, click on the "Edit (Text)" or "Edit (GUI)" link to edit the appropriate information.

System Email Management

You need to be a "WikiEditor" to be able to even read the pages indicated here. Only a "WikiAdmin" can make changes.

The [Email Management](#) page gives instructions to make changes related to the numerous automated emails that Papio regularly sends out.

Others

There are many other tasks that should probably get added here.

Appendices

Appendix 1: Excel Dates (Mac versus PC)

“1904” Date button – This option seems to cause many problems when transferring Excel files back and forth between Mac’s and PC’s. Dates are sometimes off by 4 years (1462 Days). The date system needs to be the same on both computers for this to not happen

Mac’s – Preferences- Calculations- Make sure “1904” date button IS NOT checked

PC’s – Tools – Options – Calculations – Make sure “1904” date button IS NOT checked

Appendix 2: PLHD Project, Creating tables

Creation of table Biography and Fertility for the Primate Life Histories Database (PLHD) from Babase.

S.C. Alberts, P.O. Onyango

20 Oct 2011

Column definitions revised by Jake Gordon, 18 Mar 2014

Susan and collaborators are creating a database which combines the work of 7 different long term primate studies for one master database. Our study is Study Number 2. We are producing two tables. One table (Biography) has a wide range of variables for demographic analysis. Some of these are exactly equivalent to variables in babase. Others are different. The other table (Fertility) indicates, for each female, whether there are gaps in observation large enough that a birth could have occurred without our detection. In the case of the baboons, there are none. For this reason, the construction of Fertility is equivalent to identifying the beginning and end of observations for each female.

This document is divided into two sections. The first lists the variables in the Biography table. The second provides the queries used to produce the latest update of the baboon data, for both the Biography and Fertility tables.

Part 1. Variables in Biography.

1a. The following are individuals and groups that should be excluded:

- Exclude all individuals who have never been recorded in a study group (grp 3.0 or higher)
- Exclude all cases where bstatus = 9
- Exclude all cases where sname = null or sex = U (this excludes all pregnancies that did not result in live births)
- Exclude all individuals whose statdates are earlier than 1971-07-01

1b. Below is a brief overview of the columns that should appear in the Biography file when you are finished. All dates should be in the format dd-mm-yyyy. 'Our' as used in this document means 'as exists/used in BABASE'.

- Study ID is 2 for every individual
- AnimID is equal to our sname
- AnimName is equal to our name
- Birthgroup is equal to our matgrp
- Birthgroupcertainty = C in all cases (certain; uncertainty is built into our database via birth group 9, unknown)
- Sex is equal to our sex
- Momid is equal to the first 3 characters of our pid
- Firstborn is conditional:
 - U for those individuals with a pid of XXX101 and above, or for those individuals lacking a pid
 - N for those who have a pid of XXX2 and above (and less than XXX101)
 - Y for those who have a pid of XXX1
- Birthdate is equal to our birth
- BDmin is equal to our earliestbirth
- BDmax is equal to our latestbirth
- Entrydate is the entrydate in the ENTRYDATES table (in babase_pending)
- Entrytype is the entrytype in the ENTRYDATES table
- Departdate is equal to our statdate
- Departdateerror, for our purposes, is > 0 only when the number of days between Departdate and date last seen is >15 days (it is different for other studies where animals might be seen less predictably). In practice, departdateerror is:
 - 0 for females
 - 0 for males with statdates prior to 1994 (because census date are not complete prior to 1994)
 - For males with statdates > 1 Jan 1994, the number of days between the last known census date and the statdate, expressed as a fraction of a year
- Departtype is conditional:
 - D (death) for those individuals with a biograph.status = 1.
 - P (permanent disappearance) for those who have a biograph.status > 1. (including those individuals who were in group 1.3 at the end of observation period of Proton's group in 1993, all males who disperse without a known death date, etc).
 - O (ongoing) for those individuals with biograph.status = 0

Part 2. Documentation for retrieving data for PLHD from BABASE

By Jake Gordon, Mar 2014

The queries below will add new tables to the "plhd_tables" schema. The name of the new table (set by you, the user) is **highlighted** below (in this example, the query will create a table called "plhd_2013_bio" in the plhd_tables schema). Be sure to change this text to your desired table name before executing the query.

Biography

```
WITH special_dates AS (select sq.sname, sq.memb_entry, sq2.last_census from (select sname, min(date) as memb_entry from members where grp<2.9 group by sname) sq left join (select sname, max(date) as last_census from census where status<>'A' group by sname) sq2 on sq.sname=sq2.sname)
SELECT CAST(2 as integer) as studyid, b.sname as animid, b.name as animname, b.matgrp as birthgroup, CAST('C' as varchar) as birthgroupcertainty, b.sex, SUBSTRING(b.pid FROM 1 FOR 3) as momid, CASE WHEN SUBSTRING(b.pid from 4 for 3)::smallint>=101 THEN 'U' WHEN SUBSTRING(b.pid from 4 for 3)::smallint>=2 THEN 'N' WHEN SUBSTRING(b.pid from 4 for 3)::smallint=1 THEN 'Y' ELSE 'U' END AS firstborn, b.birth as birthdate, CASE WHEN b.bstatus = 0 THEN b.birth WHEN b.bstatus = 1 THEN (b.birth - 183) WHEN b.bstatus = 2 THEN (b.birth - 365) WHEN b.bstatus = 3 THEN (b.birth - 548) WHEN b.bstatus = 4 THEN (b.birth - 730) END AS bmin, CASE WHEN b.bstatus = 0 THEN b.birth WHEN b.bstatus = 1 THEN (b.birth + 183) WHEN b.bstatus = 2 THEN (b.birth + 365) WHEN b.bstatus = 3 THEN (b.birth + 548) WHEN b.bstatus = 4 THEN (b.birth + 730) END AS bmax, CASE WHEN b.matgrp=3 THEN b.birth ELSE COALESCE(ed.entrydate, special_dates.memb_entry) END as entrydate, CASE WHEN b.matgrp>3.0 THEN 'I' WHEN b.matgrp = 1.0 AND b.birth <='1971-08-01' THEN 'O' WHEN b.matgrp = 2.0 AND b.birth <='1980-04-02' THEN 'O' ELSE 'B' END AS entrytype, b.statdate as departdate, CASE WHEN b.sex='F' THEN 0 WHEN b.statdate<'1994-01-01' THEN 0 ELSE ((b.statdate-special_dates.last_census)/365.25) END::numeric(4,3) AS departdateerror, CASE WHEN b.status=1 THEN 'D' WHEN b.status>1 THEN 'P' WHEN b.status=0 THEN 'O' END AS departtype
INTO plhd_tables.plhd_2013_bio
FROM special_dates LEFT JOIN biograph b ON special_dates.sname=b.sname LEFT JOIN babase_pending.entrydates ed ON special_dates.sname=ed.sname WHERE b.bstatus<>9 AND b.sex <>'U' AND b.statdate>'1971-07-01'
```

When exporting the table to an Excel file, reformat dates so that they're of the format, dd-mm-yyyy, and replace NULL ("N") cells with empty cells.

Fertility

As before, use the below query. In the **green-highlighted** area, indicate the name you'd like your new table to be called in the PLHD_TABLES schema. In the **yellow-highlighted** area, use the name of your biography table—whatever name you used above—but take care not to remove the "pb" at the end of that line. This query allows for some revisions of the departdate and type for special cases; for example, individuals in dropped groups that we still watch occasionally. To add manual changes, enter manual adjustments in the "DROPPED_GRPS" table in the PLHD_TABLES schema. The query below will first refer to DROPPED_GRPS, then generate fertility information for all other females based on the Biography table.

```
SELECT CAST(2 as integer) as studyid
, pb.animid
, pb.entrydate as startdate
, pb.entrytype as starttype
, --For females whose stopdates needed to be determined manually, we've set up a separate table,
DROPPED_GRPS. This section pulls their manually-set stopdates.
COALESCE (dg.stopdate, pb.departdate) as stopdate
, CASE
--If stopdate was adjusted above, stoptype needs to be adjusted to 'P'
WHEN dg.stopdate IS NOT NULL THEN 'P'
ELSE pb.departtype
END as stoptype
INTO plhd_tables.plhd_2013_fer
FROM plhd_2013_bio pb
LEFT JOIN dropped_grps dg
ON pb.animid=dg.sname
WHERE pb.sex = 'F'
```

ORDER BY pb.entrydate

As before, when exporting the table to an Excel file, reformat dates so that they're of the format, dd-mmm-yyyy, and replace NULL ("/N") cells with empty cells.

Appendix 3: Feedback to Field Team

By Patrick Onyango October 2011

We generate summaries of various data sets in order to address two main objectives 1) to help us get a sense of the rate of data collection for various ad lib data sets, 2) to identify and provide feedback to field team on errors they make when collecting the various data sets.

We briefly describe below the SQL queries used to retrieve the data, from BABASE, from which we generate the summaries that address the two objectives stated above.

A: Rate of data collection

Table 1: Number of consorts and number of copulations (mounts and ejaculations) corrected for number of mature females in each year

Number of consorts

```
select date_part('year', date) as year, count(act) as consorts from actor_actees where act = 'C' group by date_part('year', date) order by date_part('year', date)
```

Sum number of mounts and ejaculations

```
select date_part('year', date) as year, count(act) as copulations from actor_actees where act in('M', 'E') group by date_part('year', date) order by date_part('year', date)
```

Number of mature females (females that had matured dates < the last day of the year.

```
select count(distinct(census.sname)) from maturedates, census, biograph where census.sname=maturedates.sname and census.sname=biograph.sname and date_part('year', census.date) ='2011' and biograph.sex='F' and matured< '2011-12-31'
```

Table 2: MPis

(queries changed by Jake Gordon, 13 June 2013)

Number of MPis in each year

```
select date_part('year', date) as year, count(mpiid) from mpis group by year order by year
```

Number of individuals aged more than 1 yr (non-infants)

```
select sq.year, count(sq.year) from (select distinct census.sname,
date_part('YEAR', census.date) as year from census, biograph where
census.sname=biograph.sname and (census.date - biograph.birth)/365.25>1 and
census.date>='1999-01-01') sq group by sq.year order by sq.year
```

Number of distinct non-infants observed in MPIs (added by Jake Gordon, 13 June 2013)

```
select sq2.year, count(sq2.year) from (select distinct sq.sname,
date_part('YEAR', sq.date) as year from (select mpiid, date, actor as sname from
mpi_events where actor is not null union select mpiid, date, actee as sname from
mpi_events where actee is not null) sq join biograph on
sq.sname=biograph.sname where ((sq.date-biograph.birth)/365.25)>1) sq2 group
by sq2.year order by sq2.year
```

Table 3: Records of agonisms and groomings by each observer

Grooming

```
select date_part('year', date) as year, observer, actor_grp, count(iid) as
count_groomings from actor_actees where date_part('year', date) in (2010, 2011)
and act ='G' and observer in('JKW', 'RSM', 'SNS') group by date_part('year',
date), actor_grp, observer
```

Please note that observer identity may be different in future and so adjust as needed. Note also that before some time in 2010, “observer” is not always indicated in grooming and agonism records. So for some dates you can only examine total ad libs, not separated by observer (unless you count manually from the ad lib sheets).

Agonisms

```
select date_part('year', date) as year, observer, actor_grp, count(iid) as
count_agonisms from actor_actees where date_part('year', date) in (2010, 2011)
and act in('A', 'AS', 'OS', 'DS') and observer in('JKW', 'RSM', 'SNS') group by
date_part('year', date), actor_grp, observer
```

See note on observer identity above.

This query can be modified to provide data for shorter period. For instance, it can be modified to return data for the period Jan to May 2010 as follows:

- select date_part('year', date) as year, date_part('month', date) as month, observer, actor_grp, count(iid) as agonism_records from actor_actees where date_part('year', date) =2010 and date_part('month', date) in(1, 2, 3, 4, 5) and act in('A', 'AS', 'OS', 'DS') group by date_part('year', date), date_part('month', date), actor_grp, observer

Note again, as with grooming, that before some time in 2010, “observer” is not always indicated grooming and agonism records. So for some dates you can only examine total ad libs, not separated by observer (unless you count manually from the ad lib sheets).

B: Observer Error

We also summarize data that identifies any errors the field team made in a given period of data collection. Summaries are primarily made from data collected in Psion. We keep a raw data entry log that identifies the psion file that had errors and specifies action taken by the database manager. Additional information including examples of data summaries are contained in a folder in the Alberts Lab server/ABRP_Data Management/Feedback to team.