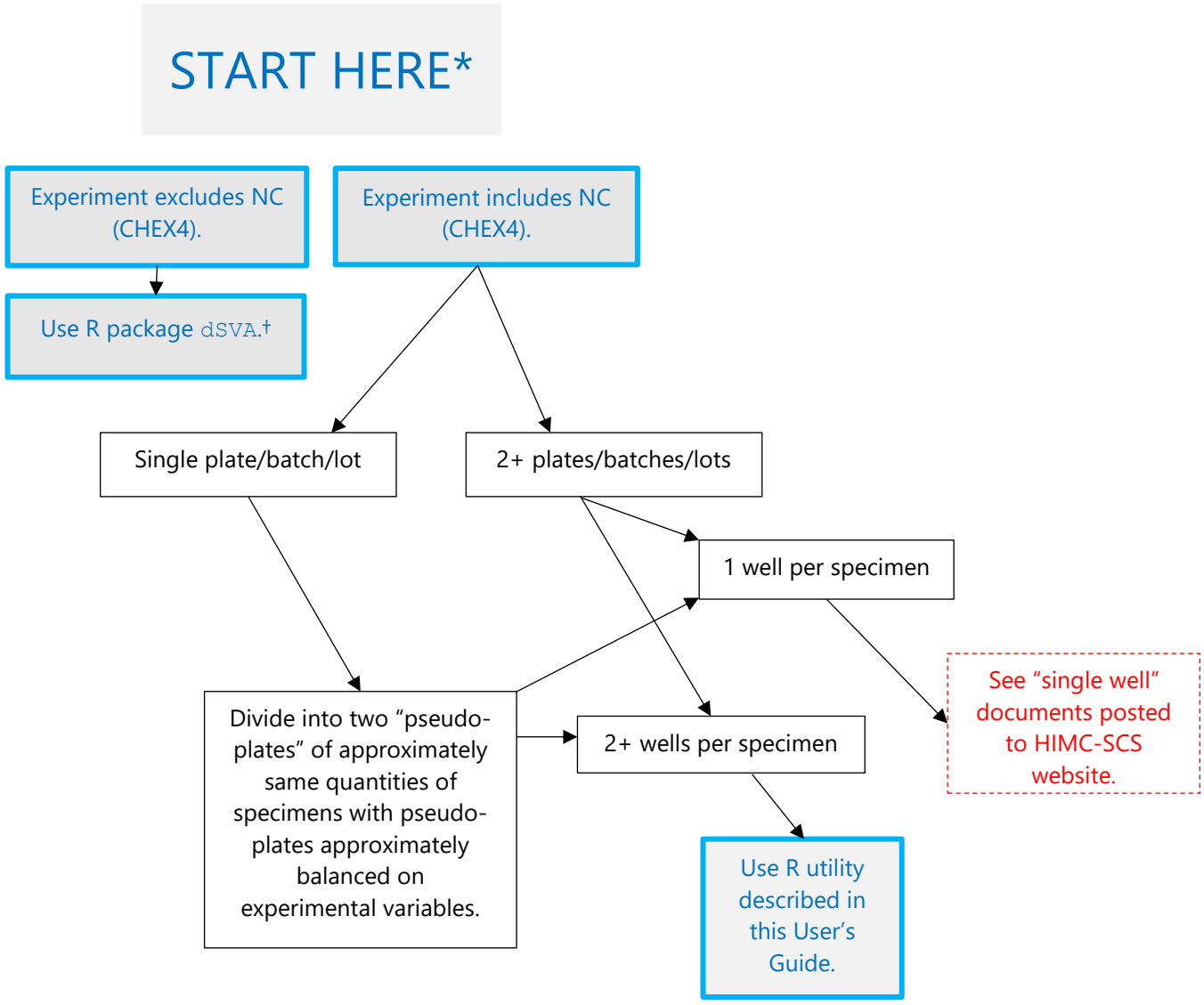

NC and CHEX4 = nonspecific binding control microbeads
MFI = median fluorescence intensity
* Contact the HIMC with any questions about this decision tree.
† Procedure suggested but currently untested by the HIMC.

Step-by-Step User's Guide for
R Utility to Correct for Plate/Batch/Lot and Nonspecific Binding Artifacts

Abbreviations:

CHEX4 = nonspecific binding control microbead
DPI = dots per inch
GUI = graphical user interface
ID = identifier
MFI = median fluorescence intensity
SP = soluble protein

Download: Utility can be obtained at http://iti.stanford.edu/himc/new-statistical-consultation-service.html.

Setting: The following assumes that you are running the utility in the R GUI supplied with base R. Details about base R language can be found in `fullrefman.pdf` that is included in base R installation.

Note: Any candidate predictor variables that might structure variation in SP MFI should be included as covariates in the utility and then all of those, except plate and CHEX4, are used as candidate predictor variables in the downstream analyses.

Step-by-Step Procedures:

1) Install most recent version of base R, appropriate for your operating system[1], obtained from www.r-project.org.

2) Install R packages `emmeans`, `lattice`, `lpme`, `nlme`, and `parallel`. Note that R language is case-sensitive. In R Console window, use command

```
install.packages(c("emmeans", "lattice", "lpme", "nlme", "parallel"))
```

[1] Utility was designed and tested in Windows 10 operating system.

3) Create input data set. First row contains variable names, where 1st column (named `Specimen`) contains specimen IDs, 2nd column (named `Plate`) contains plate/batch/lot IDs (1, 2, 3,...), 3rd column (named `Well`) contains well IDs (all unique across all rows in input dataset), 4th column (named `CHEX4`) contains CHEX4 MFI values, and remaining columns contain MFI values by SP, one column per each of the SPs, with each column named for that SP. Do not use any special characters (including blanks) in SP names (e.g., `IL17F` instead of `IL-17F`). Any remaining columns (ALWAYS to the RIGHT of the SP data) are all covariates of interest. Do not include any blank rows. The following table is an input data set (100% fake data) showing proper structure.

Specimen	Plate	Well	CHEX4	IFNG	IL10	IL21	TGFA	TGFB	VEGF	Treatment	NaUptake
1	1	1	180	278	225	232	200	415	176	Y	0.05
1	1	2	426	355	236	360	487	487	88	Y	0.01
2	1	3	310	474	157	477	36	165	237	N	0.13
2	1	4	303	73	3	59	1	315	237	N	0.14
3	2	5	320	26	154	132	277	250	307	Y	0.09
3	2	6	484	381	49	197	159	68	328	Y	0.00
4	2	7	165	449	261	60	165	283	482	N	0.08
4	2	8	458	31	179	233	291	346	452	N	0.07
5	3	9	45	195	92	488	334	173	182	Y	0.04
5	3	10	41	66	388	4	27	148	169	Y	0.16
6	3	11	144	166	495	258	181	374	443	N	0.02
6	3	12	178	195	321	126	389	305	202	N	0.10
7	4	13	145	206	77	19	459	133	220	Y	0.13
7	4	14	283	234	1	331	221	428	333	Y	0.17
8	4	15	379	278	217	439	422	73	421	N	0.13
8	4	16	415	105	435	62	332	97	310	N	0.05
9	5	17	369	413	497	386	99	29	45	Y	0.17
9	5	18	132	201	304	136	31	64	314	Y	0.10
10	5	19	453	159	229	29	460	387	484	N	0.08
10	5	20	106	281	84	180	204	206	273	N	0.14
11	6	21	75	247	466	320	151	466	257	Y	0.15
11	6	22	192	485	80	270	408	54	354	Y	0.08
12	6	23	282	323	226	134	272	139	429	N	0.10
12	6	24	231	272	124	437	196	35	303	N	0.03

Input file must be comma-delimited (e.g, `.CSV` format). File can be created in Microsoft® Excel (Microsoft® Corporation) and saved as "CSV (Comma delimited) (*.csv)."

4) In R GUI, choose "`File`" dropdown menu and choose "`New Script`". Copy and paste utility into "`R Editor`" window. Make the following assignments below the lines

```
########################
# EXAMPLE APPLICATION #
########################
```

- Supply name of directory for all output (figures and output data set), bounded by quotation marks (e.g., `Directory <- "C:\\MyData\\SP Project"`).
- Supply resolution in DPI for all output images (e.g., `Res <- 1100`).
- Supply total quantity of SPs (e.g., `QtyC <- 40`).
- Supply directory and name of input dataset, such as

```
RawData <- read.csv(file = "C:\\MyData\\SP Project\\My SP Data.csv", header = T)
```

- Supply quantity of cross-validation repeats (e.g., `Repeats <- 12`).
- Supply covariate names, separated by + sign and within quotation marks, such as

```
Covs <- "Plate + Treatment + BMI"
```

Hint:
All plates/batches/lots should be assigned to one variable named `Plate`.
Code `Plate` variable as sequential integers 1, 2, 3…

- Supply average volume (initial, pre-dilution) per specimen (e.g., `VSpecimen <- 100`).
- Supply fold dilution (e.g., `Fold <- 2`).
- Supply pipette volume (e.g., `VPipette <- 10`). Must be in same units as average volume per specimen.

5) Make changes to the `CHEX4Detrend` function, which is found below the following lines.

```
#######################
# EXAMPLE APPLICATION #
#######################
```

- Supply name of input file per assignment statement in Step 4 (e.g., `InputData = RawData`).
- Supply name of error-in-variables method. Using `Method = "HZ"` gives most recent method. See documentation for R package `lpme` for more details.
- Supply bandwidth sequence (e.g., `Seq <- 2:6`). The values in this sequence are DIVISORS; so, for example, shifting this entire sequence higher (e.g., `Seq <- 3:7`) allows the fit to be more "local," while shifting the sequence lower (e.g., `Seq <- 2:5`) can be useful when gaps appear in the sample distribution of the plate/batch/lot-detrended nonspecific binding MFI values.
- Supply column numbers of continuous covariates in input data set, such as

  ```
  Cont = c(45, 48)
  ```

  If no continuous covariates exist, use `Cont = c(0)`.
- Supply column numbers of categorical and ordinal covariates in input data set, such as

  ```
  Categ = c(46, 47, 49)
  ```

  If no categorical or ordinal covariates exist, other than Plate, use `Categ = c(0)`.

Caution:
Together, `Cont` and `Categ` should be exhaustive of all covariates
listed under `Covs <-` assignment statement of Step 4 EXCLUDING Plate.

6) Save your modified utility. Be sure that saved utility has extension `.R` (e.g., `MyUtility.R`). Close R.

7) Open R and open your modified utility in R GUI using "`Open Script`" within "`File`" dropdown menu.

8) Run utility using "`Run all`" in "`Edit`" dropdown menu in R GUI. Run may require anywhere from minutes to hours depending upon size of data set and capacity of computing environment.

Miscellaneous Notes for Troubleshooting:

1) If you wish to speed up processing, you can change the following line as follows.

```
CPUs <- floor(detectCores(all.tests = T, logical = T) / 2)
```

Beware that this increases the quantity of logical processors used by the utility and that can increase thermal stress on computer hardware.

Speed may also be boosted as follows.

```
Repeats <- 7
```

Beware that this will reduce stability of results. We have found that `Repeats <- 10` gives stable results.

As a last resort to increase speed, the following revision can be made to the script.

```
k_fold = 2
```

This setting is embedded within the `meanregbwSIMEX` function.

2) We strongly recommend that you run the utility directly within the R GUI rather than through software that runs on top of R (e.g., R Studio). Utility was not developed and tested using software that runs on top of R.

3) R package `gofortran` may be required to run the utility within Mac OS.

4) With plates as the only covariate, the `Covs` statement will be as follows.

```
Covs <- "Plate"
```

5) For best performance, completely close and re-open R between runs of the utility. Consult `fullrefman.pdf` for use of `objects` and `rm` functions for cleaning up R between runs of utility.

6) Mac users may find the following edit helpful. Everywhere in the script that starts with `png`, replace

```
res = Res, restoreConsole = T)
```

with

```
res = Res)
```

7) R package `emmeans` issues a note that "Indicator predictors are now treated as 2-level factors by default." The utility functions properly with this update to `emmeans` because indicator predictors (e.g., 0 vs. 1 or 1 vs. 2 or -1 vs. 1 coding) should be treated as 2-level categorical variables.

Further Questions?

Contact Tyson H. Holmes (tholmes@stanford.edu) regarding this R utility.

Contact Yael Rosenberg-Hasson (yaelhr@stanford.edu) regarding technical details of assays.

□