# 1. Data Flow

Grants in Original Sources
→ Grant
→ Grants
→ Summary of Grants

Demographic information
Spreadsheets and REDCap surveys
→ Summarized demographic information
→ Cohorts for analysis

# 2. Award Classifications

## Definitions

**Award-number pattern-match** = starts with #[three-character, case-insensitive pattern, {letter}##] or starts with [three-character, case-insensitive pattern, {letter}##]

## Preparation

Remove all suffixes
- -#{letter}## at end
- -#### at end
- -##{letter}# at end

Remove all prefixes for COEUS
- {letter}{letter}########{space}#{space} at beginning
- ######-###{space}#{space} at beginning

## First Pass: Lexical translation

Match based of whether string matches award number (regular expressions).

**VCTRS**: Internal K
**VCRS**: Internal K

**VPSD**: Internal K
**VFRS**: Internal K
**VA Merit**: R01 Equivalent
**VA Career**: K Equivalent
**VA CDA**: K Equivalent   // VA, not plain, CDA
**VCRS**: K12/KL2
**VCORCDP**: K12/KL2
**VEHSS**: K12/KL2
**NIEHS**: K12/KL2
VEMRT: K12/KL2
**VICMIC**: K12/KL2
**V-POCKET**: K12/KL2
**BIRCWH**: K12/KL2
**Human Frontiers in Science**: K Equivalent
**Clinical Scientist**: K Equivalent
**FTF**: K Equivalent // First Things First
**Robert Wood Johnson**: K Equivalent
**ACS**: K Equivalent
**Dermatology Foundation**: K Equivalent
**Damon Runyon Cancer Research Foundation**: K Equivalent
**LUNGevity**: K Equivalent
**AHA**: K Equivalent
**Burroughs Wellcome**: K Equivalent
**NASPGHAN**: K Equivalent   // North American Society for Pediatric Gastroenterology, Hepatology and Nutrition
**CDHNF**: K Equivalent   // Children's Digestive Health and Nutrition Foundation
**PhARMA**: K Equivalent
**NKF**: K Equivalent    // National Kidney Foundation
**SDG**: K Equivalent    // Sustainable Development Goals
**CDA**: K Equivalent    // not VA CDA (covered above)
**K Award**: Individual K
**DOD**: R01 Equivalent
**Department of Defense**: R01 Equivalent
**Peds K12**: K12/KL2

**NCI K12**: K12/KL2
**NCIK12**: K12/KL2
**NCI-K12**: K12/KL2

## Second Pass
If award number blank or 000, then N/A
If not PI and not **award-number pattern-matches** K12, then N/A
If NIH mechanism equals K12 or KL2, then K12/KL2
If **award-number pattern-matches** on K12 or KL2:
> If **award-number pattern-matches** on K12 or KL2:
>> If COEUS and PI Flag N, then do not assign.
>> If COEUS/RePORTER/ExPORTER and PI Flag Y, then Training Grant Admin
>> Else, K12/K12
> Else, K12/K12

If **award-number pattern-matches** on VUMC, then N/A
If **award-number pattern-matches** Unknown individual, then K Equivalent
If **award-number pattern-matches** R01, then R01
If NIH mechanism equals R01, then R01
If **award-number pattern-matches** R00 or K99, then K99/R00
If NIH mechanism equals R00, then K99/R00
If **award-number pattern-matches** T##, then Training Grant Admin
If budget >= 250,000 per year[1], project end – project start > 3 years, and not R01 and **award-number pattern-match** doesn't match K##, then R01-equivalent

## Third Pass
If matches K23[space]-[space], then Individual K
If **award-number pattern-matches** K24, then N/A
If **award-number pattern-matches** R03, then N/A
If **award-number pattern-matches** IK2BX or IK2CX (VA Grants), then K Equivalent

If **award-number pattern-matches** I01BX or I01CX (VA Grants), then R01 Equivalent
If **award-number pattern-matches** R37, then R01 Equivalent
If award number matches Internal K, then Internal K
If award number matches Individual K, then Individual K
If award number matches K12/KL2, then K12/KL2
If **award-number pattern-matches** K99, then K99/R00
If **award-number pattern-matches** K## and doesn't match K99, then Individual K
If NIH mechanism equals K##, then Individual K
If award sponsor equals Veterans Administration, Tennessee, then K Equivalent
If sponsor type equals Non-Profit Foundations/ Associations AND percent effort >= 50% AND direct costs >= $50,000/year for at least three years, then K Equivalent

## Final Pass
All remaining N/A

# 3. Filtering Grants
Grants are filtered from an exclude list of people whose names are similar to names in the membership. E.g., Jane Johnson might be on the exclude list for Jane Johns.

# 4. Combining Grants

Grants are also combined if they have the same **base award number** (cf. section on Calculating the Base Award Number). All grants with type of N/A are not considered in the process of combining. This process creates one grant out of a sequence of grants and makes:

---

[1] Approximately 90% (91.8% of the top 15,000 listed in FY2017) of R01 grants nationally in the NIH RePORTER meet this threshold.

- Its start to be the earliest starting date in the sequence;
- Its end to be the latest ending date in the sequence;
- Its direct budget to be combined for all of the grants in the sequence;
- Its budget to be combined for all of the grants in the sequence.

# 5. Duplicate Grants

## Definition
Two-or-more grants are considered duplicates of each other if one of the following is true:
1. They have the same **base award number** (cf. section on Calculating the Base Award Number). The script attempts to combine them (cf. section on Combining Grants).
2. They start on the same date and have the same type.

## Preference
When two-or-more grants are duplicates of each other, the following order of priority of order of sources is used, with the most-preferred source coming first:
- modify = Grant Wrangler
- coeus = COEUS (only if have data before 2008)
- exporter = NIH ExPORTER
- reporter = Federal RePORTER
- custom = Custom Grant (REDCap form)
- followup = Follow-up Survey (later surveys preferred over earlier surveys)
- scholars = Initial Initial Survey
- data = Spreadsheet Newman "data"
- sheet2 = Spreadsheet Newman "Sheet2"
- new2017 = Spreadsheet of new scholars for 2017

## Algorithm
1. Filter out exclusion list. This is a list of names which are often mistaken for members of this database. Currently, there is only one name: Harold L. Moses, the father of Harold Moses, Jr., one of our scholars.
2. Organize by data source (prioritized by order in the section Preference) and combine grants into a list, with one grant per award number. Combining grants will take place according to the protocol in the section of Combining Grants.
3. Import Grant Wrangling list of changes.
4. Make those changes to the grant list. When a grant is removed, it affects all grants with the same Base Award Number.
5. Order the grants by start date.
6. Remove duplicate base-award-numbers grouped by sources and combine these grants (grouped by sources) into a list, with one grant per base award number. Again, combining grants will take place according to the protocol in the section of Combining Grants.
7. All grants with type N/A are filtered out.
8. Remove duplicate base-award-numbers grouped by starting timestamp. If two grants start on the same date and have the same type, remove the grant that is of a less-preferred source. Again, the list in the Preference section is consulted.
9. Move data into final data structure for access. Prepare to save a copy in REDCap.

# 6. Calculating the Base Award Number
If the number can be broken up, Activity Code + Institute Code + Serial Number (usually six digits; sometimes more); otherwise, full sponsor award number.

HHS grants take the form of HHS[CHARACTER]########[CHARACTER] or HHS[CHARACTER]##########[CHARACTER].

# 7. Handling Publications

Two NCBI Databases are accessed via their API E-Utilities. For more information on these, please visit https://www.ncbi.nlm.nih.gov/books/NBK25501/.

PubMed IDs are gathered from VUMC's internal publication script (which handles self-verification by each scholar) and directly from PubMed's eSearch API. PubMed eSearch is queried via the scholar's name(s) and institution(s). *Citations without a PubMed ID are not handled.*

Information about each citation is acquired from the eFetch API with PubMed. The downloaded fields consist of: PubMed ID (PMID), PubMed Central ID (PMCID), DOI, Authors, Title, Publication Types, MESH Terms, Journal, Volume, Issue, Year, Month, Day, and Pages.

Each PubMed ID is used to access the citation on the iCite API. For more information on iCite, please visit https://icite.od.nih.gov/.

iCite provides the following fields: Is research?, Number of Citations by Other Papers, Number of Citations per Year, Expected Number of Citations per Year, Field Citation Rate, NIH Percentile, and Relative Citation Ratio. If no iCite record is matched, these fields remain blank.

There does not yet exist a foolproof way of matching a scholar's name with a particular scholar (i.e., the disambiguation problem). The publications are filtered through the known institutions, so mis-matches should be kept to a minimum. Nonetheless, each citation needs to be verified from a human.

VUMC's internal script has each scholar verify her/his matched citations. Therefore, these can be automatically included into our counts. PubMed matches that are not pre-matched need to be handled by the **Publication Wrangler**. Again, only names need to be verified in this step and nothing else.

# 7. Construction of Variables

## Grant Information

| Variable Name | Description | Data Source | Source File | Construction | Code Value / Format Label |
|---|---|---|---|---|---|
| | These are intermediate values in the code that are used for calculation. As such, they are passed to the final variables. | | | All grants are downloaded into separate forms in the REDCap record. They are then imported into a generic format that is described here. Then they are compared to each other, assembled in order of starting date, and filtered for duplicates by base award number. | |

| Sponsor award number | The full award number from the sponsor (e.g., 1R019999999-0101) | COEUS: Sponsor Award Number RePORTER: Project Number ExPORTER: Full Project Num Scholars Survey: Grant number Followup Survey: Grant number Custom Grant: Grant number Prior: Summary Award Sponsor Number Newman spreadsheets: <ul><li>If of type K12/KL2, *K12/KL2*</li><li>If assigned a sponsor type, use this value</li><li>Otherwise label *Internal K – Rec. ###* or *Unknwon Individual Rec. ###* or *Individual K – Rec. ###* or *R01* depending on placement in spreadsheet</li></ul> | Grant class | Directly assigned from data source | String |
|---|---|---|---|---|---|

| Base award number | The basic award number without specification of the support year, other suffixes, or the application type fields (e.g., R019999999) | From Grant's Sponsor Award Number | Grant class | Calculated from sponsor award number | String Correlates with Sponsor award number |
|---|---|---|---|---|---|
| Is Federal | A Boolean value whether the source is a known Federal source or not | COEUS: Direct Sponsor Type COEUS: Prime Sponsor Type ExPORTER form RePORTER form | Grant class | Perform the calculation for the direct_dollar amount, but only for grants designated Federal. <br>• If COEUS's either direct sponsor type or prime sponsor type as DOD, NASA, ED, NSF, Federal, DOE, NIH, or PHS, then Federal. <br>• If source is exporter or reporter, then Federal. <br>(defined in Grant::isFederal) | Boolean |
| Type | Award classification (bin) that we place this award in | Prior: Summary Award Type Custom: Manually Entered as Type Remaining: Calculated after all Grant information is entered | Grant class | Cf. Award Classifications (in section 2) | • 99 = N/A (not in awards list) <br>• 1 = Internal K <br>• 2 = K12/KL2 <br>• 3 = Individual K <br>• 4 = K Equivalent <br>• 5 = R01 <br>• 6 - R01 Equivalent <br>• 7 = Research Fellowship <br>• 8 = Training Grant Admin |

| Start | The starting date of the grant or of this portion of the grant (if divided into years) | Initial Survey: Grant Start<br>Followup Survey: Grant Start<br>Custom Grant: Grant Start<br>COEUS: Project Start Date<br>RePORTER: Project Start Date<br>ExPORTER: Project Start<br>Prior: Summary Award (Start) Date<br>Newman: Spreadsheet Field | Grant class | Reformatted into YYYY-MM-DD format and assigned. In Coeus, if PI Flag = N, then assigned blank. In the algorithm, these numbers can be combined with other grants with the same Base Award Number and with the same Source. | Date data (YYYY-MM-DD) |
|---|---|---|---|---|---|
| End | The ending date of the grant or of this portion of the grant (if divided into years) | Initial Survey: Grant End<br>Followup Survey: Grant End<br>Custom Grant: Grant End<br>COEUS: Project End Date<br>RePORTER: Project End Date<br>ExPORTER: Project End<br>Prior: Summary Award End Date | Grant class | Reformatted into YYYY-MM-DD format and assigned. If from Newman spreadsheets, assigned blank.<br>In the algorithm, these numbers can be combined with other grants with the same Base Award Number and with the same Source. | Date data (YYYY-MM-DD) |

| Source | The source of the data. | Prior: Summary Award Source Remaining: Hard-Coded into Grant Factories for each type of REDCap form | Grant class | Directly assigned from the form on which the data source sits | In order of priority: <br>• modify = Data Wrangler <br>• custom = Custom Grant (REDCap form) <br>• coeus = COEUS <br>• exporter = NIH ExPORTER <br>• reporter = Federal RePORTER <br>• followup = Follow-up Survey (later surveys preferred over earlier surveys) <br>• scholars = Initial Initial Survey <br>• data = Spreadsheet Newman "data" <br>• sheet2 = Spreadsheet Newman "Sheet2" <br>• new2017 = Spreadsheet of new scholars for 2017 |
|---|---|---|---|---|---|

| Budget | The total budget for this portion of the grant. Cost from dates Start to End. | Scholars Survey: calculated from Grant Costs<br>Followup Survey: calculated from Grant Costs<br>Custom Grant: calculated from Grant Costs<br>COEUS: Total Cost Budget Period<br>RePORTER: Total Cost Amount<br>ExPORTER: Total Cost<br>Prior: Summary Award Total Budget | Grant class | Directly assigned from data source or calculated from direct costs and F&A rate (Scholars Surveys, Followup Surveys, Custom Grants only), depending on how the respondent entered the budget. | Floating point |
|---|---|---|---|---|---|
| Direct Budget | The part of the grant that is available for use (after F&A is subtracted). Cost from dates Start to End. | Scholars Survey: Grant Costs<br>Followup Survey: Grant Costs<br>Custom Grant: Grant Costs<br>COEUS: Direct Cost Budget in Period from Start to End<br>RePORTER: calculated from Total Cost Amount<br>ExPORTER: Direct Cost Amt | Grant class | If RePORTER data, calculated from the F&A rate and the total budget.<br>Otherwise, directly assigned from data source. | Floating point |

| Sponsor | The organization sponsoring the grant | Scholars Survey: Organization Followup Survey: Organization Custom Grant: Organization COEUS: Direct Sponsor Name RePORTER: Agency ExPORTER: IC Name Prior: Summary Award Direct Budget | Grant class | Directly assigned from data source | String |
|---|---|---|---|---|---|
| PI Flag | A Boolean value to tell whether the scholar is a PI/Co-PI of the given grant. | Scholars Survey: Role Followup Survey: Role Custom Grant: Role COEUS: PI Flag Prior: Y | Grant class | For COEUS, directly assigned from data source. For Scholars Survey, Followup Survey, or Custom Grant, Y if PI or Co-PI or not specified (blank); otherwise, N. | String Boolean Y = PI or Co-PI N = not PI and not Co-PI |
| Prime Sponsor Type | The COEUS type of sponsor; used to determine whether a grant is Federal or not. | COEUS: Prime Sponsor Type | Grant class | If COEUS, directly assigned from the data source. Otherwise blank. | String |
| Direct Sponsor Type | The COEUS type of sponsor; used to determine whether a grant is Federal or not. | COEUS: Direct Sponsor Type | Grant class | If COEUS, directly assigned from the data source. Otherwise blank. | String |

| Sponsor Type | The type of organization from which the grant originates. | RePORTER: Agency ExPORTER: IC Name COEUS: Direct Sponsor Type Newman: Calculated Type | Grant class | From COEUS, entered directly from encoded COEUS field. For RePORTER and ExPORTER, use the organization name For Newman, calculated from splitting a spreadsheet field | String (not enumeration) |
|---|---|---|---|---|---|
| Last Update | The date that the data was last updated according to REDCap scripts. | COEUS: Last Update RePORTER: Last Update Custom Grant: Last Update ExPORTER: Last Update | Grant class | Directly assigned from the PHP script that downloads the data | Date data (YYYY-MM-DD) |
| Link | An HTML-encoded hyperlink to the page with the original information | Hard-coded for each data source | Grant class | Directly assigned from the data source | String |
| Percent Effort | The percent effort that the individual devotes to the project. | COEUS: Percent Effort Prior: Summary Award Percent Effort | Grant class | Directly assigned from data source if extant; otherwise, blank. | Integer (percentage) |
| **Summarized Awards** | | | | | |
| [order of first 15 grants] | | Grant Start, Grant Sponsor Number, Grant Base Number, Grant Source | | | |

| Date | Starting date for the award | Grant Start | Grant class | Assigned from grant data | Date data |
|---|---|---|---|---|---|
| End date | Ending date for the award | Grant End | Grant class | Assigned from grant data<br>If grant is of type "Internal K" or "K12/KL2" and no end date exists. then apply the following logic:<br>• If the scholar converts to an R01 or R01-Equivalent, the grant ends the day before the R01/R01Equiv grant begins.<br>• Otherwise, the grant ends three years after the start date. | Date data |
| Title | Title for the award | Grant Title | Grant class | Assigned from grant data | String |
| Last Update | The date of the last download (or update) for the award from the original source | Grant Last Update | Grant class | Assigned from grant data | Date data |

| Type | Award classification (bin) that we place this award in | Grant Type | Grant class | Cf. Award Classifications (in section 2) | • 99 = N/A (not in awards list)<br>• 1 = Internal K<br>• 2 = K12/KL2<br>• 3 = Individual K<br>• 4 = K Equivalent<br>• 5 = R01<br>• 6 = R01 Equivalent<br>• 7 = Research Fellowship<br>• 8 = Training Grant Admin |
|------|------|------|------|------|------|
| Source | Data source of award | Grant Source | Grant class | Calculated in 6b_makeSummary | In order of priority:<br>• modify = Data Wrangler<br>• custom = Custom Grant (REDCap form)<br>• coeus = COEUS<br>• exporter = NIH ExPORTER<br>• reporter = Federal RePORTER<br>• followup = Follow-up Survey (later surveys preferred over earlier surveys)<br>• scholars = Initial Initial Survey<br>• data = Spreadsheet Newman "data"<br>• sheet2 = Spreadsheet Newman "Sheet2"<br>• new2017 = Spreadsheet of new scholars for 2017 |

| Source type | Type of data source of award | Grant Source | | Calculated in 6b_makeSummary from source | 0 = Computer Generated<br>1 = Self-Reported<br>2 = Manually Entered |
|---|---|---|---|---|---|
| Sponsor no | Award sponsor number | Grant Sponsor Number | Grant class | Assigned from grant data | String data |
| Age | Age at start of grant (if date-of-birth specified) | Grant Start, Date of Birth | Grant class | Calculated from [start] date and identifier_dob | Floating-point number |
| NIH Mechanism | Three-letter NIH mechanism (e.g., K24, R01; if specified) | Grant Mechanism | Grant class | Directly assigned from COEUS. | String data |
| Total Budget | The total (direct + indirect) budget. Only applicable for the timespan from [start] date to end_date. | Grant Budget (total) | Grant class | Uses Grants class method to adjust for F&A.<br>= [direct budget] * (1 + F&A) unless total budget is available from source, in which case that number is used. | Floating-point number; in dollars |
| Direct Budget | The direct budget transferred after F&A adjustment. Only applicable for the timespan from [start] date to end_date. | Grant Direct Budget | Grant class | Uses Grants class method to adjust for F&A.<br>= [total budget] / (1 + F&A) unless total budget is available from source, in which case that number is used. | Floating-point number; in dollars |
| Percent Effort | Percent effort (if specified) | Grant Percent Effort | Grant class | Assigned from grant data | Integer |

## Conversion Variables

| First Any K | The date of the start of the earliest grant of type Internal K, K12/KL2, Individual K, or K Equivalent | Grant Start, Grant Type | Grants class | In the Award List, find the first (earliest) grant with the given type; save starting date here. | Date data |
|---|---|---|---|---|---|
| First Any K source | Data source of award used in first_any_k | Grant Source | Grants class | Copied from source of relevant grant. | • modify = Data Wrangler<br>• custom = Custom Grant (REDCap form)<br>• coeus = COEUS<br>• exporter = NIH ExPORTER<br>• reporter = Federal RePORTER<br>• followup = Follow-up Survey (later surveys preferred over earlier surveys)<br>• scholars = Initial Initial Survey<br>• data = Spreadsheet Newman "data"<br>• sheet2 = Spreadsheet Newman "Sheet2"<br>• new2017 = Spreadsheet of new scholars for 2017 |
| First Any K source type | Type of data source of award used in first_any_k | Grant Source | Grants class | Calculated from source of relevant grant. | 0 = Computer Generated<br>1 = Self-Reported<br>2 = Manually Entered |

| Last Any K | The date of the start of the latest grant of type Internal K, K12/KL2, Individual K, or K Equivalent | Grant Start, Grant Type | Grants class | In the Award List, find the last (latest) grant with the given type; save starting date here. | Date data |
|---|---|---|---|---|---|
| Last Any K source | Data source of award used in last_any_k | Grant Source | Grants class | Copied from source of relevant grant. | See first_any_k_source. |
| Last Any K source type | Type of data source of award used in last_any_k_source | Grant Source | Grants class | Calculated from source of relevant grant. | 0 = Computer Generated 1 = Self-Reported 2 = Manually Entered |
| First External K | The date of the start of the earliest grant of type Individual K, or K Equivalent | Grant Start, Grant Type | Grants class | In the Award List, find the first (earliest) grant with the given type; save starting date here. | Date data |
| First External K source | Data source of award used in first_external_k | Grant Source | Grants class | Copied from source of relevant grant. | See first_any_k_source. |
| First External K source type | Type of data source of award used in first_external_k_source | Grant Source | Grants class | Calculated from source of relevant grant. | 0 = Computer Generated 1 = Self-Reported 2 = Manually Entered |

| | | | | | |
|---|---|---|---|---|---|
| Last External K | The date of the start of the latest grant of type Individual K, or K Equivalent | Grant Start, Grant Type | Grants class | In the Award List, find the last (latest) grant with the given type; save starting date here. | Date data |
| Last External K source | Data source of award used in last_external_k | Grant Source | Grants class | Copied from source of relevant grant. | See first_any_k_source. |
| Last External K source type | Type of data source of award used in ast_external_k_sou rce | Grant Source | Grants class | Calculated from source of relevant grant. | 0 = Computer Generated 1 = Self-Reported 2 = Manually Entered |
| First R01 (or equivalent) | The date of the start of the earliest grant of type R01 or R01 Equivalent | Grant Start, Grant Type | Grants class | In the Award List, find the first (earliest) grant with the given type; save starting date here. | Date data |
| First R01 (or equivalent) award Type | The type of the earliest grant of type R01 or R01 Equivalent | Grant Type | Grants class | In the Award List, find the first (earliest) grant with the given type; save starting date here. | Date data |
| First R01 (or equivalent) source | Data source of award used in first_r01 | Grant Source | Grants class | Copied from source of relevant grant. | See first_any_k_source. |
| First R01 (or equivalent) source type | Type of data source of award used in last_external_k | Grant Source | Grants class | Calculated from source of relevant grant. | 0 = Computer Generated 1 = Self-Reported 2 = Manually Entered |

| **"Ever" Variables** | | | | | |
|---|---|---|---|---|---|
| Internal K | Yes/No question based on whether the scholar has ever had a grant of type Internal K. | Grant Type | Grants class | Yes if ever had any grant in the Award List categorized as type Internal K. Otherwise, No. | 1 = Yes 0 = No |
| Individual K or Equiv | Yes/No question based on whether the scholar has ever had a grant of type Individual K. | Grant Type | Grants class | Yes if ever had any grant in the Award List categorized as type Individual K or K Equivalent. Otherwise, No. | 1 = Yes 0 = No |
| K12 KL2 | Yes/No question based on whether the scholar has ever had a grant of type K12/KL2. | Grant Type | Grants class | Yes if ever had any grant in the Award List categorized as type K12/KL2. Otherwise, No. | 1 = Yes 0 = No |
| R01 or Equiv | Yes/No question based on whether the scholar has ever had a grant of type R01 or R01 Equivalent. | Grant Type | Grants class | Yes if ever had any grant in the Award List categorized as type R01 or R01 Equivalent. Otherwise, No. | 1 = Yes 0 = No |

| Last Any K to R01 or Equiv | Tells nature of K→R conversion where the K being measured is the last Internal K, K12/KL2, Individual K, or K equivalent (more common). Grouped into one of six bins, specified on right. | Grant Type, Grant Start | Grants class | *Values are described in decoded language in the table cell to the right.* Value is "1" if the last K is an Individual K or a K Equivalent and if the time between the last K and the first R01/R01-Equivalent is <= 5 years. Value is "1" if the last K is an Internal K or a K12/KL2 and if the time between the first K and the last R01/R01-Equivalent is <= 3 years. Value is "2" if the last K is an Individual K or a K Equivalent and if the time between the last K and the first R01/R01-Equivalent is > 5 years. Value is "2" if the last K is an Internal K or a K12/KL2 and if the time between the first K and the last R01/R01-Equivalent is > 3 years. Value is "3" if there is no R01/R01-Equivalent present in the list and if the last K is an Individual K or a K Equivalent and if the time between the last K and the present time is <= 5 years. Value is "3" if there is no R01/R01-Equivalent present in the list and if the last K is an Internal K or a K12/KL2 and if | Groups of bins with encoding (1-4): 1. Converted K to R01-or-Equivalent in While on K 2. Converted K to R01-or-Equivalent Not While on K 3. Still on K, No R01-or-Equivalent 4. Not on K, No R01-or-Equivalent |

| | | | | the time between the last K and the present time is <= 3 years. Value is "4" if there is no R01/R01-Equivalent present in the list and if the last K is an Individual K or a K Equivalent and if the time between the last K and the present time is > 5 years. Value is "4" if there is no R01/R01-Equivalent present in the list and if the last K is an Internal K or a K12/KL2 and if the time between the last K and the present time is > 3 years. | |
|---|---|---|---|---|---|

| (First) External K to R01 or Equiv | Tells nature of K→R conversion where the K being measured is the first Individual K or K Equivalent. Grouped into one of six bins, specified on right. | Grant Type, Grant Start | Grants class | *Values are described in decoded language in the table cell to the right.* Value is "1" if the first K is an Individual K or a K Equivalent and if the time between the first K and the first R01/R01-Equivalent is <= 5 years. Value is "2" if the first K is an Individual K or a K Equivalent and if the time between the first K and the first R01/R01-Equivalent is > 5 years. Value is "3" if there is no R01/R01-Equivalent present in the list and if the first K is an Individual K or a K Equivalent and if the time between the first K and the present time is <= 5 years. Value is "4" if there is no R01/R01-Equivalent present in the list and if the first K is an Individual K or a K Equivalent and if the time between the first K and the present time is > 5 years. | Groups of bins with encoding (1-4): 1. Converted K to R01-or-Equivalent in While on K a. R follows K within life of K award 2. Converted K to R01-or-Equivalent Not While on K a. R follows K sometime after K award expires 3. Still on K, No R01-or-Equivalent a. K award is still b. There is no R yet awarded 4. Not on K, No R01-or-Equivalent a. K award is not active b. There is no R yet awarded |
|---|---|---|---|---|---|

| Last External K to R01 or Equiv | Tells nature of K→R conversion where the K being measured is the last Individual K or K equivalent (more common). Grouped into one of six bins, specified on right. | Grant Type, Grant Start | Grants class | *Values are described in decoded language in the table cell to the right.* Value is "1" if the last K is an Individual K or a K Equivalent and if the time between the last K and the first R01/R01-Equivalent is <= 5 years. Value is "2" if the last K is an Individual K or a K Equivalent and if the time between the last K and the first R01/R01-Equivalent is > 5 years. Value is "3" if there is no R01/R01-Equivalent present in the list and if the last K is an Individual K or a K Equivalent and if the time between the last K and the present time is <= 5 years. Value is "4" if there is no R01/R01-Equivalent present in the list and if the last K is an Individual K or a K Equivalent and if the time between the last K and the present time is > 5 years. | Groups of bins with encoding (1-4): 1. Converted K to R01-or-Equivalent in While on K 2. Converted K to R01-or-Equivalent Not While on K 3. Still on K, No R01-or-Equivalent 4. Not on K, No R01-or-Equivalent |
|---|---|---|---|---|---|

| First Any K to R01 or Equiv | Tells nature of K→R conversion where the K being measured is the first Internal K, K12/KL2, Individual K, or K equivalent. Grouped into one of six bins, specified on right. | Grant Type, Grant Start | Grants class | *Values are described in decoded language in the table cell to the right.* Value is "1" if the first K is an Individual K or a K Equivalent and if the time between the first K and the first R01/R01-Equivalent is <= 5 years. Value is "1" if the first K is an Internal K or a K12/KL2 and if the time between the first K and the first R01/R01-Equivalent is <= 3 years. Value is "2" if the first K is an Individual K or a K Equivalent and if the time between the first K and the first R01/R01-Equivalent is > 5 years. Value is "2" if the first K is an Internal K or a K12/KL2 and if the time between the first K and the first R01/R01-Equivalent is > 3 years. Value is "3" if there is no R01/R01-Equivalent present in the list and if the first K is an Individual K or a K Equivalent and if the time between the first K and the present time is <= 5 years. Value is "3" if there is no R01/R01-Equivalent present in the list and if the first K is an Internal K or a K12/KL2 and if | Groups of bins with encoding (1-4): 1. Converted K to R01-or-Equivalent in While on K 2. Converted K to R01-or-Equivalent Not While on K 3. Still on K, No R01-or-Equivalent 4. Not on K, No R01-or-Equivalent |

| | | | | the time between the first K and the present time is <= 3 years. Value is "4" if there is no R01/R01-Equivalent present in the list and if the first K is an Individual K or a K Equivalent and if the time between the first K and the present time is > 5 years. Value is "4" if there is no R01/R01-Equivalent present in the list and if the first K is an Internal K or a K12/KL2 and if the time between the first K and the present time is > 3 years. | |
|---|---|---|---|---|---|

## Grants/Awards Yearly Summaries ("Summary of Grants")

| Year | The calendar year under inspection, from January 1 until December 31 | Grant Start, Grant End | Summary Grants class | One year/form-instance exists for every year represented from the start time of grant 1 to the end time of the last grant in the Award List | Integer |
|---|---|---|---|---|---|
| Total Dollar | The total (direct + indirect) dollar amount for the given year | Grant Total Budget | Summary Grants class | For each grant in Award List, <ol><li>Find all grants with same base award number.</li><li>Filter step 1's grants to include only the most-preferred source (from most-preferred to least preferred: modify, custom, coeus, exporter, reporter, followup [later surveys preferred], scholars, data, sheet2, new2017)</li><li>Calculate the fraction of each grant in the given calendar year (linearly).</li><li>Sum the total_budget-of-the-grant * fraction-in-the-calendar-year</li></ol> | Integer |

| Direct Dollar | The direct dollar amount for the given year | Grant Direct Budget | Summary Grants class | For each grant in Award List,<br>1. Find all grants with same base award number.<br>2. Filter step 1's grants to include only the most-preferred source (from most-preferred to least preferred: modify, custom, coeus, exporter, reporter, followup [later surveys preferred], scholars, data, sheet2, new2017)<br>3. Calculate the fraction of each grant in the given calendar year (linearly).<br>4. Sum the direct_budget-of-the-grant * fraction-in-the-calendar-year | Integer |
|---|---|---|---|---|---|
| Internal K or K12/KL2 | The direct dollar amount for the given year for all grants of type Internal K or K12/KL2 | Grant Direct Budget, Grant Type | Summary Grants class | Perform the calculation for the direct_dollar amount, but only for grants of type Internal K or K12/KL2 | Integer |
| Individual-K/K-Equiv | The direct dollar amount for the given year for all grants of type Individual K or K Equivalent | Grant Direct Budget, Grant Type | Summary Grants class | Perform the calculation for the direct_dollar amount, but only for grants of type Individual K or K Equivalent | Integer |

| R01/R01-Equiv | The direct dollar amount for the given year for all grants of type R01 or R01 Equivalent | Grant Direct Budget, Grant Type | Summary Grants class | Perform the calculation for the direct_dollar amount, but only for grants of type R01 or R01 Equivalent | Integer |
|---|---|---|---|---|---|
| Federal | The direct dollar amount for the given year for all grants from a known Federal source | Grant Direct Sponsor Type, Grant Prime Sponsor Type, Grant Source | Summary Grants class | Perform the calculation for the direct_dollar amount, but only for grants designated Federal.<br>• If COEUS's either direct sponsor type or prime sponsor type as DOD, NASA, ED, NSF, Federal, DOE, NIH, or PHS, then Federal.<br>• If source is exporter or reporter, then Federal.<br>• (defined in Grant::isFederal) | Integer |
| Non-federal | The direct dollar amount for the given year for all grants from a non-Federal source | Grant Direct Sponsor Type, Grant Prime Sponsor Type, Grant Source | Summary Grants class | Perform the calculation for the direct_dollar amount, but only for grants NOT designated Federal (as shown in federal) | Integer |
| | | | | | |

## Demographic Information

| Current Institution | The current institution of the scholar | 1. Manual input on the Identifiers form<br>2. Initial Import<br>3. Initial Survey | Scholar class | Vanderbilt if specified so in the Initial Survey<br>Meharry if specified so in the Initial Survey<br>Else if specified in Initial Survey Other Institution field, copy from there<br>Else fill with "Unknown" | String |
|---|---|---|---|---|---|

| Current Job Title | The job title of the scholar's current job | 1. Position Change form<br>2. Initial Survey<br>3. Followup Survey | Scholar class | Select the variable with the latest date | String |
|---|---|---|---|---|---|
| Current Job Category | The type of job that the scholar is currently involved with | 1. Position Change form<br>2. Initial Survey<br>3. Followup Survey | Scholar class | Select the variable with the latest date | 1 = Academia, still research-dominant (PI)<br>5 = Academia, still research-dominant (Staff)<br>2 = Academia, not research dominant<br>7 = Academia, training program<br>3 = Private practice<br>4 = Industry, federal, non-profit, or other - research dominant<br>6 = Industry, federal, non-profit, or other - not research dominant |
| Current Department | The name of the department, if applicable, of the current job | Position Change form | Scholar class | Select the variable with the latest date | String |
| Current Division | The current division of the scholar | Followup Survey<br>Initial Survey | Scholar class | Copy variable from the first source with data | String |

| Current Rank | The current academic rank of the scholar | | Scholar class | | 1 = Research Fellow<br>2 = Clinical Fellow<br>3 = Instructor<br>4 = Research Assistant Professor<br>5 = Assistant Professor<br>6 = Associate Professor<br>7 = Professor<br>8 = Other<br>9 = Pre-doctoral Candidate<br>10 = Post-doctoral Candidate |
|---|---|---|---|---|---|
| Under-represented minority group member | Yes/No | | Scholar class | Must have a disability, have disadvantaged status, or be a member of a specific race/ethnicity group (Black/non-Hispanic, White/Hispanic, Black/Hispanic, Asian, Other, Black/ethnicity not specified, American Indian or Native Alaskan, Native Hawaiian or Other Pacific Islander | Blank = unknown<br>0 = No<br>1 = Yes |
| Current Appointment Start | The starting date of the current academic appointment in Current Rank | | Scholar class | | |
| Tenure Status | The tenure status of the scholar | | Scholar class | | |

| Degrees | The degrees that the scholar has received | 1. Override<br>2. Followup Survey<br>3. Initial Survey<br>4. VFRS<br>5. Spreadsheet data | Scholar class | Copy variable from the first source with data | 1 = MD only<br>2 = MD + PhD<br>3 = MD + other<br>4 = PhD<br>6 = Other |
|---|---|---|---|---|---|
| Gender and Gender (Source) | The gender of the scholar | 1. Override<br>2. Initial Survey<br>3. VFRS<br>4. Initial Import<br>5. Spreadsheet data | Scholar class | Copy variable from the first source with data | 1 = Female<br>2 = Male |
| Mentor and Mentor (Source) | The scholar's primary mentor | 1. Override<br>2. Followup Survey<br>3. Initial Survey<br>4. VFRS<br>5. Spreadsheet data | Scholar class | Copy variable from the first source with data | String |
| Race/Ethnicity and Race (Source) and Ethnicity (Source) | The race and ethnicity of the scholar | 1. Override<br>2. Initial Survey<br>3. VFRS<br>4. Initial Import<br>5. Spreadsheet data | Scholar class | For both race and ethnicity, from the first source with data, separately take data and combine them into this field's categories | 1 =, White, non-Hispanic<br>2 = Black, non-Hispanic<br>3 = White, Hispanic<br>4 = Black, Hispanic<br>5 = Asian<br>6 = Other<br>7 = White, ethnicity not specified<br>8 = Black, ethnicity not specified<br>9 = American Indian or Native Alaskan<br>10 = Native Hawaiian or Other Pacific Islander |

| Date of Birth and Date of Birth (Source) | The date of birth of the scholar | 1. Initial Survey<br>2. VFRS<br>3. Initial Import<br>4. Spreadsheet data | Scholar class | Copy variable from the first source with data | Date data |
|---|---|---|---|---|---|

| Primary Department and Primary Department Source | The primary affiliated department of the scholar | 1. Override<br>2. Initial Survey<br>3. VFRS<br>4. Spreadsheet data | Scholar class | Copy variable from the first source with data | 104300 = Anesthesiology [104300]<br>104250 = Biochemistry [104250]<br>120450 = Biological Sciences [120450]<br>104785 = Biomedical Informatics [104785]<br>104286 = Cancer Biology [104286]<br>104280 = Cell and Developmental Biology [104280]<br>104226 = Center for Human Genetics Research [104226]<br>120430 = Chemistry [120430]<br>104791 = Emergency Medicine/Administration [104791]<br>104625 = Health Policy [104625]<br>104782 = Hearing And Speech Sciences [104782]<br>104216 = Institute for Global Health [104216]<br>130100 = Kennedy Center Institute (MC) [130100]<br>122450 = Mechanical Engineering [122450]<br>104368 = Medicine [104368]<br>104383 = Medicine/Allergy Pulmonary & Critical Care [104383] |

| | | | | | 104333 = Medicine/Cardiovascular Medicine [104333] |
|---|---|---|---|---|---|
| | | | | | 104342 = Medicine/Clinical Pharmacology [104342] |
| | | | | | 104348 = Medicine/Dermatology [104348] |
| | | | | | 104351 = Medicine/Diabetes Endocrinology [104351] |
| | | | | | 104370 = Medicine/Epidemiology [104370] |
| | | | | | 104355 = Medicine/Gastroenterology [104355] |
| | | | | | 104366 = Medicine/General Internal Medicine [104366] |
| | | | | | 104353 = Medicine/Genetic Medicine [104353] |
| | | | | | 104379 = Medicine/Hematology Oncology [104379] |
| | | | | | 104362 = Medicine/Infectious Disease [104362] |
| | | | | | 104375 = Medicine/Nephrology [104375] |
| | | | | | 104386 = Medicine/Rheumatology [104386] |
| | | | | | 104336 = Medicine/Stahlman Cardio Research [104336] |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | 104270 = Molecular Physiology & Biophysics [104270]<br>104400 = Neurology [104400]<br>104407 = Neurology/Cognitive Disorders [104407]<br>104403 = Neurology/Epilepsy [104403]<br>104412 = Neurology/Immunology [104412]<br>104409 = Neurology/Movement Disorders [104409]<br>104415 = Neurology/Neuromuscular [104415]<br>104418 = Neurology/Oncology [104418]<br>104410 = Neurology/Sleep Disorders [104410]<br>104425 = Obstetrics and Gynecology [104425]<br>104450 = Ophthalmology [104450]<br>104481 = Ortho - Oncology [104481]<br>104475 = Orthopaedics and Rehabilitation [104475]<br>999999 = Other (999999) |

| | | | | | 104781 = Otolaryngology [104781]<br>104500 = Pathology [104500]<br>104555 = Pediatrics/Adolescent Medicine [104555]<br>104565 = Pediatrics/Cardiology [104565]<br>104570 = Pediatrics/Child Development [104570]<br>104568 = Pediatrics/Clinical Research Office [104568]<br>104578 = Pediatrics/Critical Care Medicine [104578]<br>104582 = Pediatrics/Emergency Medicine [104582]<br>104580 = Pediatrics/Endocrinology [104580]<br>104585 = Pediatrics/Gastroenterology [104585]<br>104595 = Pediatrics/General Pediatrics [104595]<br>104590 = Pediatrics/Genetics [104590]<br>104598 = Pediatrics/Hematology [104598] |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| | | | | | 104623 = Pediatrics/Hospital Medicine [104623]<br>104606 = Pediatrics/Infectious Disease [104606]<br>104610 = Pediatrics/Neonatology [104610]<br>104600 = Pediatrics/Neurology [104600]<br>104621 = Pediatrics/Pulmonary [104621]<br>104592 = Pediatrics/Vanderbilt-Meharry Center in Sickle Cell [104592]<br>104290 = Pharmacology [104290]<br>104291 = Pharmacology/Clin Pharm [104291]<br>104795 = Physical Medicine and Rehabilitation [104795]<br>104529 = Psychiatry/Adult Psychiatry [104529]<br>104535 = Psychiatry/Child & Adolescent Psychiatry [104535]<br>120660 = Psychology [120660]<br>104675 = Radiation Oncology [104675] |

| | | | | | 104650 = Radiology and Radiological Science [104650]<br>106052 = School of Nursing - Research Faculty [106052]<br>104703 = Section of Surgical Science [104703]<br>SFS = Service Free Stipends [SFS]<br>126230 = Special Education [126230]<br>104477 = Sports Medicine [104477]<br>104705 = Surgery [104705]<br>104714 = Surgery/Liver Transplant [104714]<br>104760 = Surgery/Pediatric Surgery [104760]<br>104709 = Surgery/Surgical Oncology [104709]<br>104726 = Surgery/Thoracic Surgery [104726]<br>104717 = Surgery/Trauma [104717]<br>104775 = Urologic Surgery [104775]<br>104201 = Vanderbilt Vaccine Center [104201]<br>104268 = Biostatistics (104268)<br>104267 = Biostatistics/Cancer Biostatistics (104267) |
| --- | --- | --- | --- | --- | --- |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | 104202 = Center for Biomedical Ethics and Society (104202)<br>104790 = Emergency Medicine (104790)<br>104204 = Institute of Medicine and Public Health (104204)<br>120727 = Medicine, Health & Society (120727) |

| Left Vanderbilt and Left Vanderbilt (Source) | The date when the scholar has left Vanderbilt | 1. Followup Surveys<br>2. Initial Survey<br>3. Spreadsheet data<br>4. Override | Scholar class | For the surveys, look for an entry with Vanderbilt or VUMC as the institution. If found, look in that entry for an end-date. For the first end-date found, save this date as the date for leaving Vanderbilt.<br>If the data exist in the spreadsheet or an override column, use that date. | Date data |