```
# overall statistics calculated by year, study data type, type of transparency measur
e
fig2_data2 %>%
  select(year_split, EmpiricalStudyData, Type, n, prop, available) %>%
  print(., n = nrow(fig2_data2))
```

```
## # A tibble: 30 × 6
## # Groups:   year_split, EmpiricalStudyData [4]
##    year_split          EmpiricalStudyData Type                 n  prop availa…¹
##    <fct>               <fct>              <fct>            <int> <dbl> <fct>
##  1 Post-RC (2018-2019) Primary            "Materials "       106  75.7 " Not a…
##  2 Post-RC (2018-2019) Primary            "Materials "        34  24.3 " Avail…
##  3 Post-RC (2018-2019) Secondary          "Materials "        43  91.5 " Not a…
##  4 Post-RC (2018-2019) Secondary          "Materials "         4   8.5 " Avail…
##  5 Pre-RC (2008-09)    Primary            "Materials "       104  85.2 " Not a…
##  6 Pre-RC (2008-09)    Primary            "Materials "        18  14.8 " Avail…
##  7 Pre-RC (2008-09)    Secondary          "Materials "        47  97.9 " Not a…
##  8 Pre-RC (2008-09)    Secondary          "Materials "         1   2.1 " Avail…
##  9 Post-RC (2018-2019) Primary            "Raw Data "          5   3.6 " Avail…
## 10 Post-RC (2018-2019) Secondary          "Raw Data "         20  42.6 " Avail…
## 11 Pre-RC (2008-09)    Primary            "Raw Data "          8   6.5 " Avail…
## 12 Pre-RC (2008-09)    Secondary          "Raw Data "         17  34   " Avail…
## 13 Post-RC (2018-2019) Primary            "Raw Data "        135  96.4 " Not a…
## 14 Post-RC (2018-2019) Secondary          "Raw Data "         27  57.4 " Not a…
## 15 Pre-RC (2008-09)    Primary            "Raw Data "        115  93.5 " Not a…
## 16 Pre-RC (2008-09)    Secondary          "Raw Data "         33  66   " Not a…
## 17 Post-RC (2018-2019) Primary            "Processed Data "    5   3.6 " Avail…
## 18 Post-RC (2018-2019) Secondary          "Processed Data "    4   8.5 " Avail…
## 19 Pre-RC (2008-09)    Primary            "Processed Data "    1   0.8 " Avail…
## 20 Pre-RC (2008-09)    Secondary          "Processed Data "    2   4   " Avail…
## 21 Post-RC (2018-2019) Primary            "Processed Data "  135  96.4 " Not a…
## 22 Post-RC (2018-2019) Secondary          "Processed Data "   43  91.5 " Not a…
## 23 Pre-RC (2008-09)    Primary            "Processed Data "  122  99.2 " Not a…
## 24 Pre-RC (2008-09)    Secondary          "Processed Data "   48  96   " Not a…
## 25 Post-RC (2018-2019) Primary            "Analysis "          1   0.7 " Avail…
```

```
## 26 Post-RC (2018-2019) Secondary          "Analysis "           2   4.3 " Avail…
## 27 Post-RC (2018-2019) Primary            "Analysis "         139  99.3 " Not a…
## 28 Post-RC (2018-2019) Secondary          "Analysis "          45  95.7 " Not a…
## 29 Pre-RC (2008-09)    Primary            "Analysis "         123 100   " Not a…
## 30 Pre-RC (2008-09)    Secondary          "Analysis "          50 100   " Not a…
## # … with abbreviated variable name ¹available
```

```
# statistics calculated by year, type of transparency measure (summed over study data type)
fig2_data2 %>%
  group_by(year_split, Type, available) %>%
  summarise(total_n = sum(n), .groups = "keep") %>%
  group_by(year_split, Type) %>%
  mutate(total_prop = total_n / sum(total_n) * 100)
```

```
## # A tibble: 15 × 5
## # Groups:   year_split, Type [8]
##    year_split          Type              available        total_n total_prop
##    <fct>               <fct>             <fct>              <int>      <dbl>
##  1 Pre-RC (2008-09)    "Materials "      " Available"          19      11.2
##  2 Pre-RC (2008-09)    "Materials "      " Not available"     151      88.8
##  3 Pre-RC (2008-09)    "Raw Data "       " Available"          25      14.5
##  4 Pre-RC (2008-09)    "Raw Data "       " Not available"     148      85.5
##  5 Pre-RC (2008-09)    "Processed Data " " Available"           3       1.73
##  6 Pre-RC (2008-09)    "Processed Data " " Not available"     170      98.3
##  7 Pre-RC (2008-09)    "Analysis "       " Not available"     173     100
##  8 Post-RC (2018-2019) "Materials "      " Available"          38      20.3
##  9 Post-RC (2018-2019) "Materials "      " Not available"     149      79.7
## 10 Post-RC (2018-2019) "Raw Data "       " Available"          25      13.4
## 11 Post-RC (2018-2019) "Raw Data "       " Not available"     162      86.6
## 12 Post-RC (2018-2019) "Processed Data " " Available"           9       4.81
## 13 Post-RC (2018-2019) "Processed Data " " Not available"     178      95.2
## 14 Post-RC (2018-2019) "Analysis "       " Available"           3       1.60
## 15 Post-RC (2018-2019) "Analysis "       " Not available"     184      98.4
```

```
# statistics calculated by study data type, transparency measured (summed over years)
fig2_data2 %>%
```

```
  group_by(EmpiricalStudyData, Type, available) %>%
  summarise(total_n = sum(n), .groups = "keep") %>%
  group_by(EmpiricalStudyData, Type) %>%
  mutate(total_prop = total_n / sum(total_n) * 100)
```

```
## # A tibble: 16 × 5
## # Groups:   EmpiricalStudyData, Type [8]
##    EmpiricalStudyData Type              available        total_n total_prop
##    <fct>              <fct>             <fct>              <int>      <dbl>
##  1 Secondary          "Materials "      " Available"           5      5.26
##  2 Secondary          "Materials "      " Not available"      90     94.7
##  3 Secondary          "Raw Data "       " Available"          37     38.1
##  4 Secondary          "Raw Data "       " Not available"      60     61.9
##  5 Secondary          "Processed Data " " Available"           6      6.19
##  6 Secondary          "Processed Data " " Not available"      91     93.8
##  7 Secondary          "Analysis "       " Available"           2      2.06
##  8 Secondary          "Analysis "       " Not available"      95     97.9
##  9 Primary            "Materials "      " Available"          52     19.8
## 10 Primary            "Materials "      " Not available"     210     80.2
## 11 Primary            "Raw Data "       " Available"          13      4.94
## 12 Primary            "Raw Data "       " Not available"     250     95.1
## 13 Primary            "Processed Data " " Available"           6      2.28
## 14 Primary            "Processed Data " " Not available"     257     97.7
## 15 Primary            "Analysis "       " Available"           1      0.380
## 16 Primary            "Analysis "       " Not available"     262     99.6
```

```
# statistics calculated by study data type (summed over transparency measures and yea
rs )
fig2_data2 %>%
  group_by(Type, available) %>%
  summarise(total_n = sum(n), .groups = "keep") %>%
  group_by(Type) %>%
  mutate(total_prop = total_n / sum(total_n) * 100)
```

```
## # A tibble: 8 × 4
## # Groups:   Type [4]
##   Type              available        total_n total_prop
```

```
##   <fct>             <fct>              <int>      <dbl>
## 1 "Materials "      " Available"          57     16.0
## 2 "Materials "      " Not available"     300     84.0
## 3 "Raw Data "       " Available"          50     13.9
## 4 "Raw Data "       " Not available"     310     86.1
## 5 "Processed Data " " Available"          12      3.33
## 6 "Processed Data " " Not available"     348     96.7
## 7 "Analysis "       " Available"           3      0.833
## 8 "Analysis "       " Not available"     357     99.2
```

```
# open access statistics after fixing coding issues on the included dataset by year
openaccessDF_used %>%
  group_by(open_access, year_split) %>%
  summarize(total_n = n(), .groups = "keep") %>%
  group_by(open_access) %>%
  mutate(open_prop = total_n / sum(total_n) * 100)
```

```
## # A tibble: 6 × 4
## # Groups:   open_access [3]
##   open_access year_split total_n open_prop
##   <chr>       <chr>        <int>     <dbl>
## 1 no          After-OS       119      40.8
## 2 no          Pre-OS         173      59.2
## 3 unknown     After-OS        21      60
## 4 unknown     Pre-OS          14      40
## 5 yes         After-OS       115      59.9
## 6 yes         Pre-OS          77      40.1
```

```
# replication statistics by year
# note that we double checked these and one was not a replication study and was misco
ded
DF_filtered_ps$Replication[DF_filtered_ps$Replication == ""] <- "There is no clear st
atement that the article reports a replication study (or studies)"
DF_filtered_ps %>%
  group_by(Replication, year_split) %>%
  summarize(total_n = n(), .groups = "keep")
```

```
## # A tibble: 4 × 3
```

```
## # Groups:   Replication, year_split [4]
##   Replication                                                  year_…¹ total_n
##   <chr>                                                        <chr>     <int>
## 1 The article claims to report a replication study (or studies)  After-…       1
## 2 The article claims to report a replication study (or studies)  Pre-OS        4
## 3 There is no clear statement that the article reports a replic… After-…     186
## 4 There is no clear statement that the article reports a replic… Pre-OS      169
## # … with abbreviated variable name ¹year_split
```

```
# coi statistics by year
DF_filtered$COI[DF_filtered$COI == ""] <- "No, there is no conflict of interest state
ment"
DF_filtered %>%
  group_by(COI, year_split) %>%
  summarize(total_n = n(), .groups = "keep") %>%
  group_by(COI) %>%
  mutate(COI_prop = total_n / sum(total_n) * 100)
```

```
## # A tibble: 4 × 4
## # Groups:   COI [2]
##   COI                                                  year_…¹ total_n COI_p…²
##   <chr>                                                <chr>     <int>   <dbl>
## 1 No, there is no conflict of interest statement       After-…     209    44.8
## 2 No, there is no conflict of interest statement       Pre-OS      258    55.2
## 3 Yes, the statement says that there is no conflict of … After-…      46    88.5
## 4 Yes, the statement says that there is no conflict of … Pre-OS        6    11.5
## # … with abbreviated variable names ¹year_split, ²COI_prop
```