Spring 2022

# Wiley Journal Package: UNL Download Activity by Subject

David C. Tyler
*University of Nebraska-Lincoln*, dtyler2@unl.edu

Casey Hoeve
*University of Nebraska-Lincoln*, achoeve@unl.edu

David Macaulay
*University of Nebraska-Lincoln*, dmacaulay2@unl.edu

Robin McClanahan
*University of Nebraska-Lincoln*, rmcclanahan1@unl.edu

# Wiley Journal Package: UNL Download Activity by Subject:

## Report One:

## Excellence in Research for Australia Field of Research Divisions, The Unsub Journal List, and the 2014–2020 COUNTER Download Data

Author:

David C. Tyler

Data:

Casey Hoeve

David Macaulay

Robin McClanahan

Submitted:

Spring 2022

**TABLE OF CONTENTS**

## FINDINGS SUMMARY

- The package as loaded into UnSub appears to have been comprised of 1,326 journals, as identified by Wiley Journal Identification Codes (JICs)
- Most journals appear to have been part of the package for all seven years of the 2014-2020 interval for which the UNL Libraries has COUNTER data
- Unsub employed a hierarchical subject classification system comprised of twenty-two subject divisions (i.e., Field of Research) and a catch-all (i.e., "Multidisciplinary") for journals that did not fit neatly into one of the other subjects (note: just over one hundred journals had no subject assigned to them, and the author tagged them "Unassigned")
- The distribution of downloads was very unequal within the package (i.e., close to the Pareto 80/20 distribution), so a sub-package comprised of journals from the $5^{th}$ and $4^{th}$ download quintiles would meet nearly all of UNL's Wiley content needs
- When the journals were grouped by subject, the distributions of downloads was also highly unequal between the subject categories, and a handful of subjects, almost all of them from the sciences and engineering, produced the bulk of the package's downloads
- If a download-quintile-based sub-package would not be agreeable to Wiley, then UNL could still meet the great bulk of its Wiley content needs through subscriptions to subject-based sub-packages instead
- Though some subject categories were not comparably productive, there likely are handfuls of high-use journals within each to which the UNL Libraries ought to consider subscribing

# INTRODUCTION

## 1) Wiley Journal Package (Unsub) and COUNTER Data

Starting in the fall of 2021, the Collection Strategies Committee (CSC) turned its attention to the Wiley journal package. As was noted in a previous report, there were numerous issues with the compilation of Wiley data, and the analysis below should therefore be considered the result of the data providers' best efforts at identifying the journals that comprise the Wiley package and at analyzing the journals' downloads by subject. Those with an interest in some of the discrepancies and/or inconsistencies that the myriad of Wiley spreadsheets presented are invited to see Appendix A of the aforementioned previous report, "Wiley Journals: UNL 2014–2020," for more information.

The data here were compiled and/or provided by Casey Hoeve, David Macaulay, and Robin MacClanahan in 2021 and then prepared for analysis by David Tyler. The data were previously employed both to look into UNL's usage of and general interest in the Wiley journal package and for test runs of the UnSub subscription analysis tool. Some of the data in those previous analyses came from COUNTER reports, and some came from UnSub.

As was noted in the report on UnSub, UnSub appears to have undercounted Wiley downloads (perhaps neglecting downloads of pre-2010 content?), so this report will be using COUNTER data. The COUNTER data, unfortunately, was not without its own issues. For example, due to changes in COUNTER standards, the 2019 and 2020 data did not include some journals present in earlier years' download reports because the COUNTER 5 standard mandates that zero-use journals not be reported. These missing journals have been included in this report via zero imputation. Also, there were titles present in the COUNTER reports that did not appear in the Wiley package as loaded into the UnSub tool. The author is are not entirely certain why these excluded titles did not appear as part of the package in Unsub, but for this report, the author will be using the package list as it appeared in UnSub since the journals not listed did not have subjects assigned to them. Additionally, instead of using the journals' titles to identify them, the author will be making use of Wiley's Journal Identification Codes (JICs) because these JICs seemed to remain fairly consistent across title changes within the datasets. For instance, if a journal were to change its title three times over the interval, its data would appear as that of three separate titles in the COUNTER reports but would be listed as belonging to just one JIC (albeit, with three separate titles attached).

The result of all of this fiddling with the data is that this report will analyze 1,326 JICs with 1,085,455 total COUNTER downloads during the 2014-2020 interval. The report will neglect the 770 JICs, with 122,281 total COUNTER downloads over the interval, that were not part of the Wiley package as it appeared in UnSub. Therefore, 36.7% of the JICs that matched journal titles in COUNTER download reports for Wiley and 10.1% of the COUNTER-recorded downloads for Wiley will not be part of this analysis. If these JICs were part of the Wiley package and should have been included in UnSub, then a substantial part of the long, lower-use tail of the Wiley download distribution will be missing from this analysis. If these missing JICs were not part of the Wiley package but were subscribed to, then perhaps they ought to be

analyzed separately, for the UNL Libraries may be spending subscription dollars on a sizeable number of lower-use journals.

**2) Note on Reading the Report**

Before proceeding, the author would like, here, quickly to provide a further brief note of caution to the reader:

First, the data and analysis provided here should be taken to reflect UNL's usage of and/or "revealed preferences" in the Wiley Journal Package at the *institutional* level. The reader should not infer too much about the preferences and interests of individual departments and programs at UNL (i.e., not commit the ecological fallacy). The data were collected at the level of the institution, and the author has no data on the journals' users. The author, therefore, cannot calculate a rate of usage for departments, programs, or types of patrons. Therefore, it would be inappropriate to look that the data for, say, Biological Sciences (315,517 downloads) and Studies in Creative Arts and Writing (600 downloads) and incautiously conclude that Biological Sciences loves its Wiley journals, but Creative Arts and Writing does not. The difference in measured activity may mask a confound or two (e.g., the programs may be of radically different sizes), so differences in download activity may be, at least in part, a measure of how many library patrons various departments and programs have rather than a pure measure of how avid a Wiley user the various departments and programs are. So, this report should be employed as a first step in the analysis of the package by subject, rather than as the last word.

Second, throughout, the reader should keep the Pareto (80/20) distribution in the back of her/his mind while reading this report and should assume, speaking roughly, that it is present fractally at most, if not all, levels of measurement. That is to say, one should think that roughly 80% of the downloads for the package were produced by roughly 20% of the journals, roughly 80% of the downloads within each subject sub-set were produced by roughly 20% of the subjects' journals, and so on. This pattern should be more present the more journals there are within a measured group (e.g., it should be very clearly present in the package [1,326 journals], it should be pretty clearly present in *06-Biological Sciences* [200], but it might be pretty inchoate in *19-Studies in Creative Arts and Writing* [5]). Just because the 80/20 pattern is not entirely obvious in the smaller subject groups, one should not rush to conclude that they are different. The safer conclusion would be cautiously to assume that, were more journals added to the smaller subjects, the distribution of their downloads would increasingly come to resemble the 80/20 distributions of the larger subject groups. This cannot be assumed with absolute certainty (e.g., a 60/40 distribution might exist), but it is the safer assumption and should pertain in the preponderance of cases.

So, one should expect that a small portion of each group, at all levels of measurement, will have produced the bulk of the measured activity. One should not jump to conclude that all journals in high-use subjects are high use nor that all of the journals in low-use subjects exhibit low use. There should be a sizable number of low-use journals in high-use subjects, and there probably will be a handful of relatively higher-use journals in most low-use subjects.

## INFORMAL ANALYSIS

### 3) The Journal Subjects (ERA Field of Research Divisions)

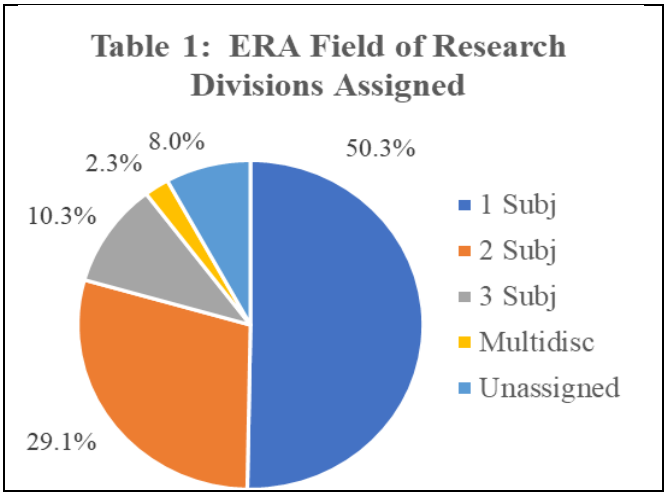
UnSub assigned subjects to most of the journals in the Wiley package using the Excellence in Research for Australia (ERA) system of the Australian Research Council. Information about the ERA subjects is available online here:

https://www.arc.gov.au/excellence-research-australia/era-2018-journal-list

So far as the author can ascertain from a quick glance, the ERA system has a hierarchical structure and assigns "Field of Research" (FoR) codes to journals at three levels: From the examples on the site, the "Division" level would seem the broadest and most general, e.g., "31 Biological sciences"; the "Group" level provides a mid-level of specificity within each Division, e.g., "3103 Ecology"; and the "Field" level offers the greatest degree of specificity within each Group, e.g., "310301 Behavioural ecology." The system seems fairly similar to other library classification systems, such as the Library of Congress system of classifications, sub-classifications, and call numbers or the Dewey Decimal call numbering system.

UnSub employed subjects at the "Division" and "Group" levels, but not at the "Field" level, and assigned 1-3 Divisions and 1-3 Groups to most journals. For a handful of journals, UnSub also employed the unnumbered Division "MD Multidisciplinary," and another group of journals had no subjects assigned. For purposes of analysis, the author assigned the code "8888" to Multidisciplinary journals and the code "9999" to Unassigned journals.

The distribution of Division codes in the UnSub dataset was as follows in Table 1 below. As the pie chart shows, about half of the Wiley journals had just a single subject at the Division level attached to them in the UnSub dataset, just under 30% had two, just over 10% had three, just over 2% were assigned "Multidisciplinary," and 8%, for reasons unknown, did not have any ERA divisions assigned to them. ERA divisions assigned to them.



The ERA FoR Divisions employed by UnSub, with subject numbers assigned by the author, were as follows:

**Table 2: ERA Field of Research Divisions Employed by UnSub**

| Subject # | ERA Division |
|---|---|
| 1 | ['01', 'Mathematical Sciences'] |
| 2 | ['02', 'Physical Sciences'] |
| 3 | ['03', 'Chemical Sciences'] |
| 4 | ['04', 'Earth Sciences'] |
| 5 | ['05', 'Environmental Sciences'] |
| 6 | ['06', 'Biological Sciences'] |
| 7 | ['07', 'Agricultural and Veterinary Sciences'] |
| 8 | ['08', 'Information and Computing Sciences'] |
| 9 | ['09', 'Engineering'] |
| 10 | ['10', 'Technology'] |
| 11 | ['11', 'Medical and Health Sciences'] |
| 12 | ['12', 'Built Environment and Design'] |
| 13 | ['13', 'Education'] |
| 14 | ['14', 'Economics'] |
| 15 | ['15', 'Commerce, Management, Tourism and Services'] |
| 16 | ['16', 'Studies in Human Society'] |
| 17 | ['17', 'Psychology and Cognitive Sciences'] |
| 18 | ['18', 'Law and Legal Studies'] |
| 19 | ['19', 'Studies in Creative Arts and Writing'] |
| 20 | ['20', 'Language, Communication and Culture'] |
| 21 | ['21', 'History and Archaeology'] |
| 22 | ['22', 'Philosophy and Religious Studies'] |
| 8888 | [['MD', 'Multidisciplinary']] |
| 9999 | Unassigned (Missing Data) |

For this analysis, the author will only be looking at Division-level subjects. Analyses at the Group level, if desired and possible, may be undertaken on an ad hoc basis in the future. Within this report's tables, to save space, the author will largely be employing Subject #s as short names, so the reader may want to have the table above handy for reference while glancing through this report.

### 4) Downloads by Subject Area

As can be seen from the table below (Table 3), the value to UNL of the journals in a number of ERA FoR Divisions would seem obvious. For example, *02-Physical Sciences*, *03-Chemical Sciences*, *04-Earth Sciences*, *05-Environmental Sciences*, *06-Biological Sciences*, *07-Agricultural and Veterinary Sciences*, and *09-Engineering* all produced sizeable numbers of downloads over the interval (Range: 63,860 to 315,517), and these Divisions also had higher than average downloads per journal (Range: 148.2 to 375.4). If Wiley were to offer subscriptions to smaller, subject-based sub-packages, the UNL Libraries would likely subscribe to most, if not all, of these packages.

Others among the ERA FoR Divisions might be of less interest. For example, *12-Built Environment and Design* and *19-Studies in Creative Arts and Writing* produced very little

download activity over the interval, and their per journal download averages were quite low when compared to the package average.

The performances of a few Divisions may be a bit more ambiguous: *11-Medical and Health Sciences* produced a sizeable number of downloads but had a low per journal average, and *17-Psychology and Cognitive Sciences* had a somewhat similar profile, with high total downloads but only average per journal downloads. The Division *MD-Multidisciplinary*, on the contrary, had a somewhat opposite profile, with a more modest download total but a high per journal average. It is possible that usage in these subjects might stray farther from the Pareto 80/20 distribution than does the package itself, and these three subjects may warrant a closer look before UNL makes decisions concerning them.

The COUNTER data for the ERA FoR Divisions was distributed as follows:

**Table 3: ERA Field of Research Divisions: Download Performance by Subject Number**

| Subj # | JICs* | Total Dwnlds | Yearly Avg | JIC Avg |
|---|---|---|---|---|
| **Package** | **1,326** | **1,085,455** | **159,586.8** | **120.4** |
| 01 | 50 | 17,221 | 2,460.1 | 49.8 |
| 02 | 34 | 88,596 | 12,656.6 | 375.4 |
| 03 | 103 | 256,387 | 36,626.7 | 360.1 |
| 04 | 63 | 63,860 | 9,122.9 | 148.2 |
| 05 | 49 | 110,352 | 15,764.6 | 334.4 |
| 06 | 200 | 315,517 | 45,073.9 | 230.8 |
| 07 | 81 | 129,693 | 18,527.6 | 242.0 |
| 08 | 36 | 14,942 | 2,134.6 | 60.3 |
| 09 | 143 | 232,910 | 33,272.9 | 239.1 |
| 10 | 21 | 20,387 | 2,912.4 | 138.7 |
| 11 | 372 | 199,989 | 28,569.9 | 78.6 |
| 12 | 13 | 2,885 | 412.1 | 31.7 |
| 13 | 54 | 42,936 | 6,133.7 | 117.3 |
| 14 | 79 | 20,675 | 2,953.6 | 38.1 |
| 15 | 104 | 60,883 | 8,697.6 | 85.0 |
| 16 | 164 | 79,102 | 11,300.3 | 69.4 |
| 17 | 141 | 112,771 | 16,110.1 | 115.4 |
| 18 | 26 | 6,924 | 989.1 | 38.0 |
| 19 | 5 | 600 | 85.7 | 17.6 |
| 20 | 31 | 15,040 | 2,148.6 | 70.0 |
| 21 | 32 | 6,096 | 870.9 | 27.6 |
| 22 | 46 | 23,321 | 3,331.6 | 72.9 |
| 8888 | 31 | 53,579 | 7,654.1 | 248.1 |
| 9999 | 106 | 19,245 | 2,749.3 | 30.1 |

* NOTE: the ERA subjects' JIC total exceeds the count for the package (1,984 vs 1,326) because a percentage of JICs have multiple ERA FoR subjects assigned to them (see Table 1 above).

As was noted in the Introduction, some clean-up of the Wiley data was performed prior to analysis, and the reader should keep this in mind, especially where the JIC Average is concerned. The JIC Avg is, for 2014-2018 COUNTER 4 data, based on JICs with COUNTER-reported downloads. For 2019-2020 COUNTER 5 data, the JIC Avg is based on JICs with COUNTER-reported downloads and on zero-imputations for JICs missing from 2019-2020 but present from 2014-2018, the assumption being that the missing JICs were missing because they had zero

downloads in 2019 and/or 2020, rather than that they were missing because they had been removed from the package. If it were to turn out that these missing JICs actually had been dropped from the package, then these averages would have to be recalculated.

JICs missing from early reports but appearing in later reports were assumed to be additions to the package (e.g., *Journal of Operations Management* [JOOM] appears to have been added to *01-Mathematical Sciences* in 2018), so data from these JICs was included from the year of appearance forward, with no zero imputation for prior years (i.e., JOOM's data were counted for just 3 years [2018-2020]). The result will be that, for some subjects, JIC Avg will not equal Total Downloads divided by JIC count divided by total package years (7). The JIC count is for the entire interval, and the author did not convert the counts for partially present JICs into fractions. So, calculations using the Table 3 data, rather than the original data, will produce some small discrepancies here and there. Barring future adjustments to the table due to canceled/dropped journals, the JIC Avg as reported should be treated as the best download productivity estimate that the UNL Libraries has for each ERA FoR Divisions' journals.

The main takeaway from Table 3, though, should be that, within the universe of the Wiley package, there are a number of subject areas in whose journals UNL has shown an obvious interest, that there are a number of subject areas in whose journals UNL has shown little interest, and that there are a few subject areas where UNL's behavior has been ambiguous and where a more in-depth analysis would probably be worthwhile, were conditions favorable to the breaking up of the package ever to arise.

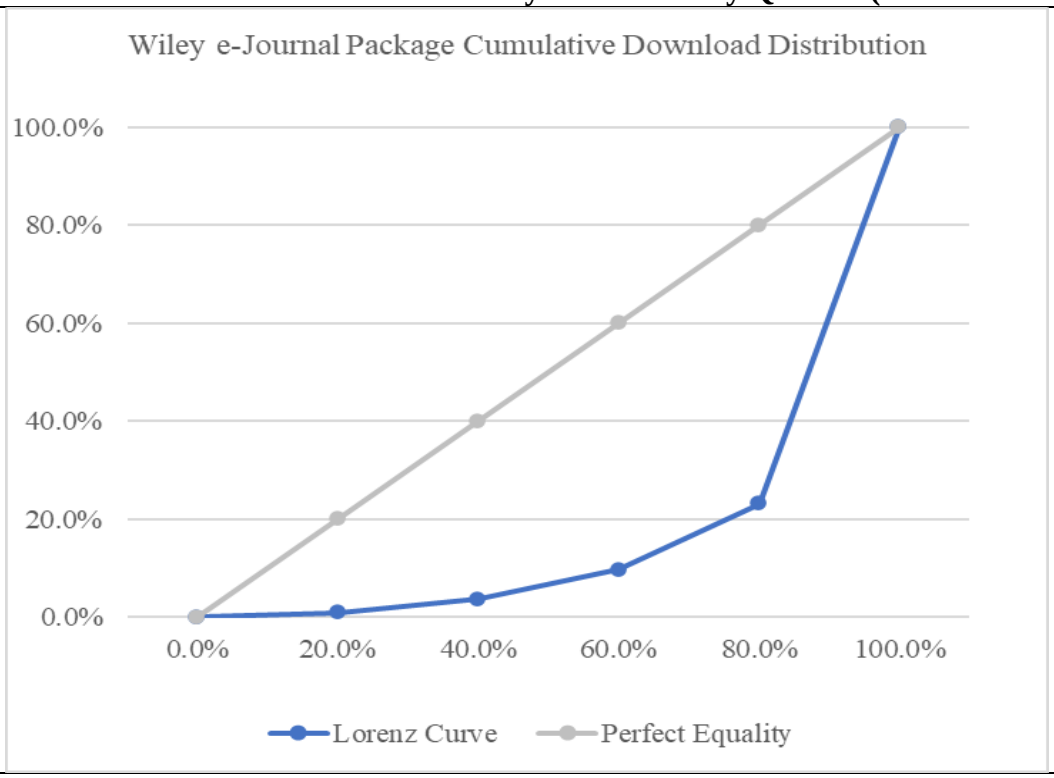
**5) Download Distributions**

Of course, Wiley may well have no interest in dividing its current journal package into subject-based sub-packages. Instead, if the UNL Libraries must pursue a different subscription strategy with Wiley, such as individual subscriptions to journals, then distribution of downloads and the stability of distribution would be of great importance. As was mentioned above, the author is inclined to assume that the distribution of downloads in the Wiley package, given its size, should resemble that of other, similar packages, such as the previously analyzed Springer package. So, a first step would be to arrange the journals and the data and see whether the Pareto 80/20 distribution is present in the Wiley package.

As can be seen in Table 4, the Wiley package exhibits the sort of unbalanced distribution of downloads that has been demonstrated by other journal packages, such as the Springer package. One can also see, however, that Wiley's distribution strays a bit from the ideal Pareto 80/20 distribution. The bottom 80% of Wiley journals produced roughly 23.2% of the interval's downloads, and the top 20% produced 76.8% of downloads. As a result, the author's grouping of the journals into quintiles may have distorted Wiley's distribution a bit, and a smoothed graph probably would place the Wiley distribution's inflection point somewhere between 80/20 and 75/25. Therefore, if the UNL Libraries were wanting to identify a group of top-producing Wiley journals by UNL downloads, the UNL Libraries might probably want to extend the group's boundary beyond the 5$^{th}$ quintile into the 4$^{th}$ quintile a wee bit. Alternately, it might be

worthwhile to graph the data by deciles or percentiles to better determine exactly where the distribution turns sharply upwards.

**Table 4: Cumulative Distribution of Wiley Downloads by Quintile (Lorenz Curve)**




Note: N = 1,326 JICs, 1,085,455 Downloads, 7 years; a small percentage of JICs were added to the collection over the course of the interval; missing 2019-2020 JICs were included via zero imputation.

The next item of interest would be the composition, by subject, of the top quintiles.

The top (5th) quintile was comprised as is shown in Table 5. As one can see by the table, a tremendous amount of the downloads for Wiley were concentrated in this 5th quintile. The quintile contained 266 journals (JICs), with 471 subjects assigned to them. Interestingly, the standard deviation for the journals' total downloads is more than 1.5 times as great as the journals download average, which suggests that the performances of the journals were quite variable, a suggestion supported by the very wide minimum-maximum range.

**Table 5: Composition of the 5th Download Quintile of Wiley Journals**

| Variables | Statistics |
|---|---|
| Journal Identification Codes (JICs) | 266 |
| Field of Research Divisions (FoRs) | 22* |
| FoRs Assigned | 471* |
| Total Downloads (7 years) | 834,271 |
| Average Downloads | 3,136.4 |
| Standard Deviation | 5,062.6 |
| Minimum | 863 |
| Maximum | 50,864 |

*Note: Includes "Unassigned" (5 JICs); 2 FoR Divisions did not appear in the 5th quintile

As one can see on the next page, the distribution of downloads by subject (i.e., Field of Research Divisions) within the 5th quintile was very unequal, and a handful of subjects were very productive for UNL within this top quintile. The subjects that appear to have been most productive, in terms of total downloads produced and in

terms of downloads per journal, were as follows: *02-Physical Sciences*, *03-Chemical Sciences*, *04-Earth Sciences*, *05-Environmental Sciences*, *06-Biological Sciences*, *07-Agricultural and Veterinary Sciences*, *09-Engineering*, and *MD-Multidisciplinary* (code: 8888). In addition to these, *11-Medical and Health Sciences* and *17-Psychological and Cognitive Sciences* produced large download totals, but their per journal download averages were more modest. The remainder of the Division codes either were much less represented within the top quintile, were much less productive within the quintile, had low per journal averages, etc.

**Table 5a: Composition of the 5th Download Quintile of Wiley Journals by ERA Field of Research Division Codes (n = 266 JICs, 471 FoRs)**

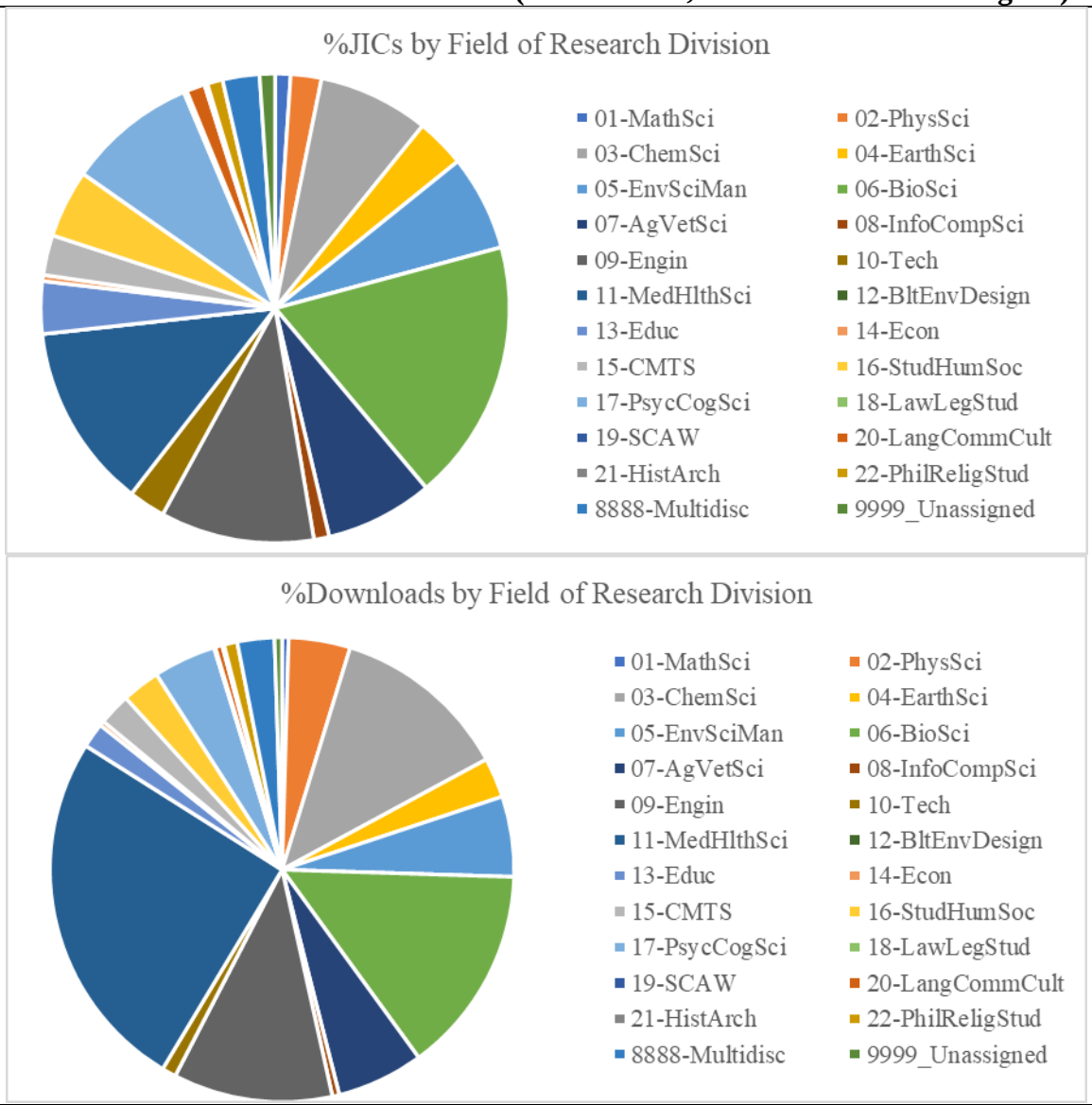
| FoRs | JICs | Dwnlds | Avg. | St. Dev. | Min. | Max. | % JICs | % Dwnlds |
|---|---|---|---|---|---|---|---|---|
| 01-MathSci | 5 | 8,421 | 1,684.2 | 347.1 | 1,325 | 2,109 | 1.1% | 1.0% |
| 02-PhysSci | 10 | 79,973 | 7,997.3 | 15,489.2 | 1,027 | 49,501 | 2.1% | 9.6% |
| 03-ChemSci | 36 | 230,611 | 6,405.9 | 11,426.4 | 924 | 50,864 | 7.6% | 27.6% |
| 04-EarthSci | 16 | 51,679 | 3,229.9 | 3,502.9 | 1,079 | 15,177 | 3.4% | 6.2% |
| 05-EnvSciMan | 31 | 104,131 | 3,359.1 | 2,771.7 | 937 | 9,834 | 6.6% | 12.5% |
| 06-BioSci | 85 | 269,883 | 3,175.1 | 3,496.3 | 902 | 22,571 | 18.0% | 32.3% |
| 07-AgVetSci | 35 | 112,687 | 3,219.6 | 3,427.9 | 902 | 18,718 | 7.4% | 13.5% |
| 08-InfoCompSci | 5 | 8,983 | 1,796.6 | 947.4 | 1,144 | 3,456 | 1.1% | 1.1% |
| 09-Engin | 50 | 206,933 | 4,138.7 | 7,451.1 | 867 | 49,501 | 10.6% | 24.8% |
| 10-Tech | 12 | 18,296 | 1,524.7 | 507.3 | 1,038 | 2,731 | 2.5% | 2.2% |
| 11-MedHlthSci | 60 | 472,041 | 7,867.4 | 9,127.7 | 885 | 9,777 | 12.7% | 56.6% |
| 12-BltEnvDesign | 0 | 0 | 0.0 | 0.0 | 0 | 0 | 0.0% | 0.0% |
| 13-Educ | 17 | 32,258 | 1,897.5 | 1,005.1 | 878 | 5,060 | 3.6% | 3.9% |
| 14-Econ | 2 | 6,091 | 3,045.5 | 1,632.7 | 1,891 | 4,200 | 0.4% | 0.7% |
| 15-CMTS | 13 | 41,412 | 3,185.5 | 3,035.0 | 920 | 11,570 | 2.8% | 5.0% |
| 16-StudHumSoc | 22 | 49,553 | 2,252.4 | 2,165.2 | 863 | 11,268 | 4.7% | 5.9% |
| 17-PsycCogSci | 42 | 81,154 | 1,932.2 | 1,884.3 | 863 | 11,268 | 8.9% | 9.7% |
| 18-LawLegStud | 1 | 1,834 | 1,834.0 | 0.0 | 1,834 | 1,834 | 0.2% | 0.2% |
| 19-SCAW | 0 | 0 | 0.0 | 0.0 | 0 | 0 | 0.0% | 0.0% |
| 20-LangCommCult | 6 | 9,063 | 1,510.5 | 504.3 | 908 | 2,285 | 1.3% | 1.1% |
| 21-HistArch | 1 | 2,434 | 2,434.0 | 0.0 | 2,434 | 2,434 | 0.2% | 0.3% |
| 22-PhilReligStud | 5 | 17,243 | 3,448.6 | 4,412.2 | 923 | 11,268 | 1.1% | 2.1% |
| 8888-Multidisc | 12 | 48,168 | 4,014.0 | 2,566.6 | 902 | 7,962 | 2.5% | 5.8% |
| 9999_Unassigned | 5 | 9,669 | 1,933.8 | 1,353.3 | 1,021 | 4,215 | 1.1% | 1.2% |

Table 5a would suggest that, if there was a cost savings to be had, a handful of subject-based mini-packages might be more attractive to UNL than the Wiley package in its entirety, with some number of top-producing journals from the less germane subject packages also being selectively picked for individual subscriptions. Whether Wiley would be open to considering such an approach at a price point and inflation rate acceptable to UNL is, of course, unknown.

Table 5a above contains quite a bit of information to take in. Readers who prefer the visual representation of data are invited to review Table 5b below, which indicates the percentage of JICs attached to each subject and the percentage of downloads attached to the same. The reader should keep in mind that there were numerous instances where multiple JICs and download counts were attached to one another. Of the 266 JICs that comprised the 5th quintile, 34.2% (91) had only a single FoR Division assigned to them, 41.7% (111) had two assigned, and 17.7% (47) had three assigned. Of the remainder, just 1.9% (5) were listed as "Multidisciplinary," and 4.5% (12) were tagged as "Unassigned" by the author. Thus, a substantial portion of the subjects

could be thought of as double-dipping, or even triple-dipping, in the pool of downloads. If Wiley were to offer its journals through single-subject mini-packages and to restrict journals to a single sub-package each, the UNL Libraries would have to untangle the data presented here in accord with the Wiley sub-package parameters offered.

**Table 5b: Percentage Composition of the 5th Download Quintile of Wiley Journals by ERA Field of Research Division Codes (n = 266 JICs, 471 FoR Divisions Assigned)**




As one can see, the FoR Divisions mentioned above as top performers in Table 5a appear as large wedges in the top and bottom pie charts of Table 5b. The top chart indicates the percentage of the package's JICs that have the indicated subjects attached to them, and the bottom chart indicates the percentage of the downloads for JICs with the indicated subjects attached to them. Because multiple subjects were attached to some JICs, there is some distortion in the charts (e.g., the bottom download chart takes 233% of downloads and reduces them to fit a 100% circle, so Biological Sciences accounts for about 32% of downloads but 14% of the pie). Still, even with

these visual distortions, one may see which were the more productive subjects. So, if Wiley was unwilling to allow the UNL Libraries to subscribe to a smaller package comprised of the top download quintile of the Wiley package but was willing to allow libraries to subscribe to smaller subject packages, the data-driven portion of UNL's subscription decisions would seem fairly obvious from the above.

Of some, but somewhat lesser, interest to UNL would be the 4th quintile. Because of how the download distribution curves, this quintile represents the bulk of UNL's remaining downloads, and a portion of the top journals of this quintile perform nearly as well as the bottom journals of the 5th quintile. If the UNL Libraries were to venture beyond the border of the 5th quintile for subscriptions, the bulk of the journals subscribed to for data-driven reasons would likely come from this 4th quintile.

**Table 6: Composition of the 4th Download Quintile of Wiley Journals**

| Variables | Statistics |
|---|---|
| Journal Identification Codes (JICs) | 265 |
| Field of Research Divisions (FoRs) | 23* |
| FoRs Assigned | 412* |
| Total Downloads (7 years) | 146,184 |
| Average Downloads | 551.6 |
| Standard Deviation | 146.8 |
| Minimum | 355 |
| Maximum | 860 |

*Note: Includes "Unassigned" (9 JICs); 1 FoR Division did not appear in the 4th quintile

As one may see from Table 6, the 4th quintile produced considerably fewer downloads than did the 5th quintile (i.e., approximate ratio 1:5.7). But these top two quintiles, when combined, would account for 90.3% on UNL's Wiley downloads over the seven-year interval. If UNL could subscribe to these two quintiles, almost the entirety of UNL's needs that could not be handled by Interlibrary Loan would be met.

The slope of the 4th quintile is considerably less steep than that of the 5th, so the Min.-Max. range was only 505 downloads. Essentially, the journals that comprise this quintile can, from top to bottom, be considered roughly equivalent in their productivity, and they are within the same order of magnitude as the bottom 14.3% of JICs in the 5th quintile. So, one could think of them as an extension of the bottom group of Wiley's top performers, or, as was noted in the previous paragraph, as journals that produce a bit more download activity (i.e., 50 to 123 downloads per year) than could be absorbed by Interlibrary Loan alone. Depending upon how much of these journals' download activity is regularly produced by fungible demand, the UNL Libraries might be able to successfully cancel some of the JICs in this quintile, but the outcome for cancelling all of them would likely not be too pleasant.

As can be seen by the discrepancy between the number of JICs and the number of Field of Research Division codes assigned in Unsub, a substantial portion of the JICs had multiple subjects assigned to them, as was the case in the 5th quintile. The bulk had a single FoR Division assigned to them, but there was still a substantial portion that were double-dipping and triple-dipping in the download pool: Of the 265 JICs that comprised the 4th quintile, 51.7% (137) had only a single FoR Division assigned, 29.8% (79) had two assigned, 12.8% (34) had three assigned, 3.4% (9) were listed as "Multidisciplinary," and just 2.3% (6) were tagged as

"Unassigned" by the author. So, the journals in the 4th quintile were more likely to have had just a single subject assigned to describe them in Unsub. Whether this would mean, for the package as a whole, that journals that have more subjects assigned to them tended to be more productive than journals with a single subject, excluding "Multidisciplinary," assigned to them is a potentially interesting question that would require formal analysis. This informal glance at the top two Wiley quintiles certainly suggests that it could be the case.

If one reviews Table 6a below, one can see that, by and large, roughly the same set of culprits produced most of the downloads of the 4th quintile. One of the differences between this table and the 5th quintile's table (Table 5a) above is that some of the 5th quintiles lesser lights were considerably more present and productive within the bounds of the 4th quintile. For example, *Education*; *Economics*; *Commerce, Management, Tourism and Service*; *Studies in Human Society*; and *Psychology and Cognitive Sciences* all produced more downloads than did *Physical Sciences* in the 4th quintile. The latter even outproduced *Chemistry*.

This in-quintile levelling of the FoR Divisions' performances is further reflected in the subjects' %JICs and %Dwnlds being more nearly equal. Put another way: after the curve of the package's download distribution became gentler beyond the 4th quintile/5th quintile inflection point, the disproportionate download performances largely disappeared. The author will not be reproducing the pie charts of Table 5b here, but one could easily imagine them in a Table 6b as being nearly identical, top and bottom.

**Table 6a: Composition of the 4th Download Quintile of Wiley Journals by ERA Field of Research Division Codes (n = 265 JICs, 412 FoR Divisions Assigned)**

| FoRs | JICs | Dwnlds | Avg. | St. Dev. | Min. | Max. | % JICs | % Dwnlds |
|---|---|---|---|---|---|---|---|---|
| 01-MathSci | 7 | 3,336 | 476.6 | 149.5 | 368 | 782 | 2.6% | 2.3% |
| 02-PhysSci | 10 | 6,163 | 616.3 | 183.0 | 419 | 857 | 3.8% | 4.2% |
| 03-ChemSci | 32 | 19,326 | 603.9 | 149.1 | 366 | 850 | 12.1% | 13.2% |
| 04-EarthSci | 11 | 5,993 | 582.8 | 163.5 | 418 | 857 | 4.2% | 4.1% |
| 05-EnvSciMan | 7 | 4,400 | 628.6 | 140.8 | 461 | 791 | 2.6% | 3.0% |
| 06-BioSci | 61 | 35,504 | 582.0 | 153.9 | 355 | 845 | 23.0% | 24.3% |
| 07-AgVetSci | 19 | 11,707 | 616.2 | 144.4 | 385 | 841 | 7.2% | 8.0% |
| 08-InfoCompSci | 4 | 2,657 | 664.3 | 139.3 | 495 | 833 | 1.5% | 1.8% |
| 09-Engin | 30 | 16,517 | 550.6 | 146.0 | 365 | 850 | 11.3% | 11.3% |
| 10-Tech | 2 | 959 | 479.5 | 103.9 | 406 | 553 | 0.8% | 0.7% |
| 11-MedHlthSci | 87 | 46,166 | 530.6 | 143.3 | 357 | 848 | 32.8% | 31.6% |
| 12-BltEnvDesign | 3 | 1,363 | 454.3 | 116.1 | 360 | 584 | 1.1% | 0.9% |
| 13-Educ | 12 | 7,416 | 618.0 | 158.7 | 358 | 835 | 4.5% | 5.1% |
| 14-Econ | 12 | 7,030 | 585.8 | 177.5 | 360 | 860 | 4.5% | 4.8% |
| 15-CMTS | 19 | 10,161 | 534.8 | 143.3 | 368 | 781 | 7.2% | 7.0% |
| 16-StudHumSoc | 28 | 14,165 | 505.9 | 146.8 | 357 | 860 | 10.6% | 9.7% |
| 17-PsycCogSci | 38 | 20,792 | 547.2 | 145.6 | 358 | 848 | 14.3% | 14.2% |
| 18-LawLegStud | 4 | 2,215 | 553.8 | 158.9 | 374 | 753 | 1.5% | 1.5% |
| 19-SCAW | 0 | 0 | 0.0 | 0.0 | 0 | 0 | 0.0% | 0.0% |
| 20-LangCommCult | 6 | 4,082 | 680.3 | 151.4 | 439 | 848 | 2.3% | 2.8% |
| 21-HistArch | 1 | 579 | 579.0 | 0.0 | 579 | 579 | 0.4% | 0.4% |
| 22-PhilReligStud | 4 | 2,177 | 544.3 | 124.9 | 378 | 681 | 1.5% | 1.5% |
| 8888-Multidisc | 6 | 3,079 | 513.2 | 113.8 | 405 | 678 | 2.3% | 2.1% |
| 9999_Unassigned | 9 | 4,964 | 551.6 | 158.2 | 366 | 807 | 3.4% | 3.4% |

If the UNL Libraries were to subscribe to the 4th quintile as a necessary complement to the 5th, the author would expect that many of the most important Wiley journals outside of the sciences and engineering would be picked up in the addition. Overall, the number of non-science Wiley journals available to UNL patrons might be greatly reduced by this strategy, but the portion of non-science downloads lost should not be all that great.

Depending upon Wiley's amenability and faculty members' stated needs and wants, some additional small number of journals from the 1st-3rd quintiles could be subscribed to on a case-by-case basis, but the UNL Libraries would be subscribing to these journals for reasons other than to meet local demand for their content.

The above, of course, has been an informal analysis based entirely on the author's perusal of the 2014-2020 Wiley COUNTER statistics for the Wiley package as defined within the Unsub subscription analysis tool. Readers with an interest in a more formal analysis of the package's subjects (i.e., ERA FoR Divisions) may turn to the next page.

## FORMAL ANALYSIS

This portion of the report will, undoubtedly, be the less enjoyable, and the author would suggest that the more casual reader eschew it. The author's years of experience in the UNL Libraries would suggest that, for most decision-making, the informal analysis above would normally be a sufficient analysis of the download data, insofar as download data might be employed to make collections decisions. So, again, this portion of the report should be of interest just to those who would like to go a bit beyond the informal analysis.

In this section, the author will attempt to address the main questions underlying the more informal portion of the report above (i.e., Do some subjects produce more downloads? If so, how big is the difference?). To approach these questions, the author will be using the same data as was used in the informal portion of the report, but he will be arranging and handling the data a bit differently. In essence, the author will be tying individual data cells to subject categories and then using statistical analysis to look into whether or not there were potentially real differences in the subject categories' download performances. This effort will be somewhat hampered by the composition of the dataset – for example, the FoR Division *19-Studies in Creative Arts and Writing* only has five journals in it, so the author will be rather limited in what he can conclude about *SCAW* – but the author will do what he can and see what can be found out.

As was the case above, there were 1,326 JICs reported in the Unsub dataset, and the UNL Libraries had seven years' worth of download counts for the Wiley package, running from 2014 through 2020. This means that there were 9,282 cells in which Wiley could have reported data. Of these, 263 were left blank because of journals' being added to the package after 2014 (Note: two journals were dropped from the package early in the interval, and the author removed them from the analysis entirely). As was mentioned above, the COUNTER rules for reporting data for 2019 and 2020 (COUNTER 5) were different than the rules employed for the 2014-2018 data, and substantial numbers of journals (255 and 284, respectively) disappeared from the COUNTER reports in these years, either because the journals had been dropped from the package or because the journals had zero downloads to report. For this analysis, the author has assumed the latter and filled in these journals' 2019 and 2020 data cells via zero imputation. Should it turn out that the missing journals actually had been dropped from the package, then this analysis will be slightly inaccurate. The result of the above fiddling will be that 9,019 cells have analyzable data.

Also, as was the case above, most of the journals in the package had subjects assigned to them (see the discussion above of Excellence in Research for Australia Field of Research Divisions) in the Unsub subscription analysis tool. Most journals had one to three of twenty-two subjects assigned to them, with a small number having been assigned "Multidisciplinary" as a sort of catch-all. About 8% of the package's journals had no subjects attached to them, and the author tagged them "Unassigned."

With this quick description of the dataset out of the way, let us move on to the questions:

**Question #1: Were there real differences in download performance by subject, and if so, how substantial were they?**

To attempt to address this major question, the author, because of the nature of the dependent (response) variable, employed the Generalized Linear Model (GLZM), a generalization of the more familiar linear regression that allows the linear model to be related to the response variable via a link function. In this case, the analytical technique employed was negative binomial (NB) regression. The more familiar General Linear Model and Ordinary Least Squares regression, for example, could not be employed at least in part because the response variable is discrete. Poisson regression, which may be the more familiar approach for dependent variables that are counts, could not be employed without an adjustment because of the Poisson model's assumption that the variance is equal to the mean. The data here appeared to be overdispersed for the Poisson model and to be very right-skewed, so the author elected to go with a model with less-restrictive assumptions (NB). The author is not a statistician at all, so attempting to fit the data to a Poisson hurdle model did not seem feasible without an additional several semesters in graduate school or some outside assistance.

On the off chance that someone more statistically capable happens to read this report and to have questions, the author has included some additional information in Table 7:

**Table 7: Generalized Linear Model: Wiley Journal Package**

Model Information

| | |
|---|---|
| Dependent Variable: | Dwnlds |
| Probability Distribution: | Negative binomial (MLE) |
| Link Function: | Log |

Case Processing Summary

| | N | Percent |
|---|---|---|
| Included | 9,019 | 97.2% |
| Excluded | 263 | 2.8% |
| Total | 9,282 | 100.0% |

Continuous Variable Information

| | |
|---|---|
| Minimum | 0 |
| Maximum | 9,976 |
| Mean | 120.35 |
| Standard Deviation | 385.254 |

Goodness of Fit

| | Value | DF | Value/DF |
|---|---|---|---|
| Deviance | 10,949.748 | 8,993 | 1.218 |
| Pearson Chi-Square | 24,437.761 | 8,993 | 2.717 |

As one can see from the table, the model is not an absolutely perfect fit for the data, but the Deviance and Chi-Square values were both low, so the model is likely sufficiently good.

This NB regression analysis will involve twenty-four categorical variables (i.e., Factors), which will be the twenty-two FoR Divisions, Multidisciplinary, and the "Unassigned" tag introduced by the author. Since most of the information about these subject categories was provided in the Informal Analysis section of the report, the author will, in the interest of saving space, omit the table detailing these variables' general characteristics here.

The first question to ask of the data is whether there is a statistically significant effect present in the data. As Table 7a indicates, there was (Note: Sig. = .000 denotes a p value < .0005):

**Table 7a: Omnibus Test**[a]

| Likelihood Ration Chi-Square | DF | Sig. |
|---|---|---|
| 3,041.035 | 24 | .000 |

a: Compares the fitted model against the intercept-only model

Having established that there was at least one statistically significant effect in the dataset, the next step would be to determine which of the independent variables (i.e., the FoR Divisions) may have produced statistically significant effects. Normally, at this point, the author would have created another table that presented the results of a test of model effects (Type III Test of Fixed Effects) to follow the Omnibus Test. This test would have shown which subjects appeared to have produced statistically significant effects via a Likelihood Ratio Chi-Square value and a significance (p) value. To save space, the author will not be producing that table here; all of the subjects, with the exception of chemistry (552.679), had values between 10,000 and 12,500, and all p values were < .0005. Thus, producing the large model effects table with all twenty-four variables listed would seem unnecessary. Suffice to say, there would appear to be some real differences in by-subject performance in the Wiley dataset.

The more interesting questions would be: What and how large were they? This question is addressed by the parameter estimates presented in Table 7b below.

As seems often to be the case, asking questions of a mountain of data seems to produce an avalanche of numbers. Such is the case here, and the resultant table can be confusing and intimidating, especially to readers unfamiliar with statistical analysis. So, a brief explanation of the table seems warranted.

- Parameter: The first column indicates the question being addressed. The first row (i.e., intercept) describes the Wiley package, and each subsequent number describes the results for a subject that is part of the package (e.g., #1 indicates 01-Mathematical Sciences; see Table 2 above for the full list of subjects).
- B: Beta indicates the slope for each question examined. A positive slope indicates that the parameter in question is associated with an increase in the dependent variable (i.e., downloads), and a negative slope indicates that the parameter is associated with a reduction in the dependent variable.
- S.E.: Standard Error is measure of sampling error (e.g., error in estimates resulting from random fluctuations in samples) and gives one a sense of how off one might be. It is

reduced by low variability and/or large samples and increased by high variability and/or small samples.

- Conf. Int.: Confidence Intervals are the lower and upper bounds of the estimate. In this case, 95% Wald Confidence Intervals were employed, which means that, were the author to collect 100 'samples' of the same size and composition as the one employed here, one would expect for the parameter estimates to be between the lower and upper bounds 95% of the time. Contrary to popular understanding, the C.I. does not necessarily mean that one is 95% confident that the parameter estimate is correct.
- Wald Chi-Square: This is a statistical test that looks into the squared ratio of the Estimate to the Standard Error of the predictor (i.e., the subject in question). It is looking into the probability that a particular test statistic as extreme as, or more so, than the one observed would occur if the null hypothesis (i.e., there is no effect or difference) were true. Large values would usually be read as indicating that it would be pretty unlikely for the observed Estimate to be observed if the null hypothesis were so.
- DF: Degrees of Freedom, which is just the number of values free to vary independently of one another when computing a statistic.
- Sig.: Significance, which indicates how likely it would be that an observed characteristic of the 'sample' would have occurred by chance. A small value (i.e., < .05) would indicate that it is unlikely that the observed effect, difference, estimate, or parameter would be produced by chance. Essentially, this value should be read as indicating whether or not what one thinks one is seeing in the data can be trusted as being real.
- Exp(B): This value is the exponent of the slope (B) and indicates the factor by which the parameter variable has an effect on the dependent variable when controlling for the effects of the other variables in the model. For example, an Exp(B) of 1.76 for a subject would indicate that the presence of that subject in a JIC's record was associated with an increase in the download rate by a factor of 1.76; a second subject with a negative slope would have an Exp(B) of less than 1.00 and would be associated with a reduction in the download rate by that factor (e.g., 01-Mathematical Sciences has a B = -.331 and an Exp(B) = .718, so having math attached to a JIC would be associated with a reduction in download rate for a typical Wiley package journal from X to .718X). In some social science, such as Education or Social Work, this value is often reported as the Incidence Rate Ratio (IRR), and it can be read as the effect of the slope on the data of the parameter in question.

So, by and large, the variables of interest in the Table 7b below would be the slope (B), which indicates whether a subject appears to add something or subtract something from the behavior of the typical Wiley package journal: the exponent of the slope (Exp(B)), which indicates how much the slope of a subject can be said to add or subtract from the behavior of the typical Wiley package journals; and the significance, which indicates whether or not what one seems to be seeing in the data is likely to be real or not. If it turns out not, such as was the case for *19-Studies in Creative Arts and Writing*, then this may be the result of their being too few respondents in the dataset, of the variability of the data for the respondents being too great, or both.

As one can see from the table, twenty of the subjects' results passed the threshold for statistical significance, so the author would be inclined to believe that there is something(s) going on in the dataset.

**Table 7b: Parameter Estimates for the Wiley Journal Package Subjects (ERA FoR Divisions): (Negative binomial regression)**

| | | | 95% Wald Conf. Int. | | Hypothesis Test | | | | 95% Wald Conf. Int. for Exp(B) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | B[a] | S.E. | Lower | Upper | Wald Chi-Square | DF | Sig. | Exp(B) | Lower | Upper |
| (Intercept) | 4.113 | .0407 | 4.033 | 4.193 | 10,203.113 | 1 | .000 | 61.123 | 56.435 | 66.201 |
| #1 | -.331 | .0755 | -.479 | -.183 | 19.235 | 1 | .000 | .718 | .620 | .833 |
| #2 | .284 | .0944 | .099 | .469 | 9.060 | 1 | .003 | 1.329 | 1.104 | 1.599 |
| #3 | 1.312 | .0600 | 1.194 | 1.430 | 477.336 | 1 | .000 | 3.713 | 3.301 | 4.177 |
| #4 | .458 | .0708 | .319 | .596 | 41.713 | 1 | .000 | 1.580 | 1.375 | 1.816 |
| #5 | .627 | .0808 | .468 | .785 | 60.110 | 1 | .000 | 1.871 | 1.597 | 2.192 |
| #6 | .962 | .0456 | .873 | 1.052 | 445.889 | 1 | .000 | 2.617 | 2.394 | 2.862 |
| #7 | .562 | .0640 | .437 | .688 | 77.017 | 1 | .000 | 1.754 | 1.547 | 1.989 |
| #8 | -.493 | .0874 | -.664 | -.322 | 31.844 | 1 | .000 | .611 | .515 | .725 |
| #9 | .720 | .0504 | .621 | .819 | 203.792 | 1 | .000 | 2.054 | 1.861 | 2.268 |
| #10 | -.058 | .1116 | -.276 | .161 | .266 | 1 | .606 | .944 | .759 | 1.175 |
| #11 | -.180 | .0414 | -.261 | -.098 | 18.843 | 1 | .000 | .836 | .771 | .906 |
| #12 | -.602 | .1440 | -.884 | -.320 | 17.496 | 1 | .000 | .548 | .413 | .726 |
| #13 | .504 | .0750 | .357 | .651 | 45.166 | 1 | .000 | 1.656 | 1.429 | 1.918 |
| #14 | -.485 | .0662 | -.615 | -.356 | 53.762 | 1 | .000 | .616 | .541 | .701 |
| #15 | .280 | .0593 | .163 | .396 | 22.211 | 1 | .000 | 1.323 | 1.177 | 1.486 |
| #16 | .118 | .0508 | .018 | .217 | 5.347 | 1 | .021 | 1.125 | 1.018 | 1.242 |
| #17 | .619 | .0480 | .525 | .713 | 166.088 | 1 | .000 | 1.857 | 1.690 | 2.040 |
| #18 | -.743 | .1024 | -.944 | -.543 | 52.737 | 1 | .000 | .475 | .389 | .581 |
| #19 | -.286 | .2435 | -.763 | .191 | 1.378 | 1 | .240 | .751 | .466 | 1.211 |
| #20 | -.182 | .0958 | -.370 | .006 | 3.596 | 1 | .058 | .834 | .691 | 1.006 |
| #21 | -1.244 | .0981 | -1.436 | -1.052 | 160.855 | 1 | .000 | .288 | .238 | .349 |
| #22 | -.100 | .0818 | -.260 | .060 | 1.500 | 1 | .221 | .905 | .771 | 1.062 |
| #8888 | 1.401 | .0973 | 1.210 | 1.591 | 207.206 | 1 | .000 | 4.058 | 3.354 | 4.911 |
| #9999 | -.708 | .0659 | -.837 | -.579 | 115.324 | 1 | .000 | .493 | .433 | .561 |
| (Scale) | 1[b] | | | | | | | | | |
| (NB) | 1.683 | .0220 | 1.641 | 1.727 | | | | | | |

Dependent Variable: Dwnlds
Model: (Intercept), #01, #02, #03, #04, #05, #06, #07, #08, #09, #10, #11, #12, #13, #14, #15, #16, #17, #18, #19, #20, #21, #22, #8888, #9999
a. Subject = 0 set to zero because this parameter is redundant
b. Fixed at the displayed value

Unsurprisingly, given the sizeable Range (0 to 9,976**)** reported in Table 7, the overall trend for the package (Intercept model) would appear to be upward. A substantial number of the subjects would appear to be associated with additions to the package's performance. Of these, a handful demonstrated sizeable performance advantages. In particular, subjects *03-Chemical Science*, *06-Biological Sciences*, *09-Engineering*, and *8888-Multidisciplinary* have Exp(B) values greater than 2.00. The value of the other subjects with Exp(B) values greater than 1.00 should not be ignored, but the author would caution that losing the top-performing journals from the listed FoR Divisions would be a real blow to the UNL departments and programs that utilize them.

On the other side of the coin, there were several subjects with negative slopes, the worst of which would appear to be *21-History and Archaeology*. Unfortunately, because of the confound inherent in the data, the author cannot confidently conclude that these FoR Divisions performed comparatively poorly because UNL in general or the pertinent departments and programs in particular did not think highly of these Wiley journals. As was noted above, relatively poor performance numbers might indicate that there were few germane patrons at UNL. The author would, however, be fairly comfortable in concluding that demand for the content of many of the journals comprising these poorly performing FoR Divisions could be met via interlibrary loan. If the UNL Libraries had to drop some subjects or subjects' journals from the Wiley package, the FoR Divisions with negative slopes would be the more attractive candidates.

This is not to say that the results are entirely unambiguous. For example, *11-Medical and Health Sciences* produced a substantial portion of the package's downloads in the informal section of the report, but here it has a slightly negative slope in the table. If the author had to guess, it may be that medicine has a large number of journals and that many of these are middling-to-poor performers. Likely, the Libraries would prefer to retain some portion of Wiley's medicine journals even if cuts must be made. Still, the journals of the medicine FoR Division likely would be far less locally important than the journals in chemistry, for example.

To assist with the review of Table 7b, the author has reproduced Table 2 here:

**Table 2: ERA Field of Research Divisions Employed by UnSub**

| Subject # | ERA Division |
|---|---|
| 1 | ['01', 'Mathematical Sciences'] |
| 2 | ['02', 'Physical Sciences'] |
| 3 | ['03', 'Chemical Sciences'] |
| 4 | ['04', 'Earth Sciences'] |
| 5 | ['05', 'Environmental Sciences'] |
| 6 | ['06', 'Biological Sciences'] |
| 7 | ['07', 'Agricultural and Veterinary Sciences'] |
| 8 | ['08', 'Information and Computing Sciences'] |
| 9 | ['09', 'Engineering'] |
| 10 | ['10', 'Technology'] |
| 11 | ['11', 'Medical and Health Sciences'] |
| 12 | ['12', 'Built Environment and Design'] |
| 13 | ['13', 'Education'] |
| 14 | ['14', 'Economics'] |
| 15 | ['15', 'Commerce, Management, Tourism and Services'] |
| 16 | ['16', 'Studies in Human Society'] |
| 17 | ['17', 'Psychology and Cognitive Sciences'] |
| 18 | ['18', 'Law and Legal Studies'] |
| 19 | ['19', 'Studies in Creative Arts and Writing'] |
| 20 | ['20', 'Language, Communication and Culture'] |
| 21 | ['21', 'History and Archaeology'] |
| 22 | ['22', 'Philosophy and Religious Studies'] |
| 8888 | [['MD', 'Multidisciplinary']] |
| 9999 | Unassigned (Missing Data) |