

The Bikard Twins FIVES Data

by

Michaël Bikard

Assistant Professor of Strategy

INSEAD

©2020

1. Conditions of Use

The data sets described in this document are part of the FIVES Project on Firm and Industry Evolution, Entrepreneurship, and Strategy, organized by Constance Helfat (Tuck School of Business at Dartmouth). If you download or otherwise obtain any of the data sets described in this document, or obtain modified versions of these data sets, you agree to abide by the conditions of use set forth here.

The organizers of the FIVES Project are not responsible for the content of any of the FIVES Project data sets. The FIVES Project organizers and the FIVES Project data set authors make no representations or warranties of any kind concerning any of the FIVES Project data, including with regard to the absence or presence of errors.

All FIVES Project data sets are protected by copyright. The copyright holders for each data set are indicated on the first page of the documentation for each data set. Permission is granted to reproduce FIVES Project data for non-profit educational and research use only. Requests to reproduce FIVES Project data for other uses should be addressed to: Professor Constance Helfat, Tuck School of Business at Dartmouth, Hanover, NH 03755, constance.helfat@dartmouth.edu.

This document describes the Bikard Twins FIVES Data, which should be referred to by this name in any derivative works. In any written and published work, users of these data should cite this document, the accompanying original article by the data set author (see below, Section 3), and the FIVES Project: Data Overview (Helfat & Klepper, 2007).

You may create a new data set by modifying a FIVES Project data set, including but not limited to adding or removing data or merging data from different FIVES Project data sets. If you utilize any FIVES Project data to create a new data set, within six years from the date that you obtain the FIVES Project data, you must provide the FIVES Project with a copy of your modified data set by contacting: Constance Helfat, constance.helfat@dartmouth.edu.

Any written documents or statistical estimates that use FIVES Project data, including but not limited to working papers and publications, must be reported to: Constance Helfat, constance.helfat@dartmouth.edu.

2. Dataset Author

Michaël Bikard

Email: michael.bikard@insead.edu

Phone: +33 1 60 72 93 80

Campus: Fontainebleau

Assistant Professor of Strategy

INSEAD

3. Dataset Description

Potential Paper Twins

The data includes 10,927 pairs of frequently co-cited papers from PubMed that have no author in common and that were published no more than 1 calendar year apart. In the case of triplets, quadruplets, etc., the same papers are listed in several pairs. The data were collected in August 2019. There is an accompanying article published in *Strategic Management Journal* by Michaël Bikard entitled “Idea Twins: Simultaneous Discoveries as a Research Tool” (doi: 10.1002/smj.3162).

The list of potential paper twins provides the following information, contained in the file

BikardTwinsFIVESdata:

- *PubMed ID* is the identifier used by PubMed for each paper. Information about each paper can be retrieved from the PubMed website: <https://www.ncbi.nlm.nih.gov/pubmed/>
- *Jaccard Index* measures the extent to which two papers are cited in the same follow-on literature. For example, a Jaccard Index of 62% means that 62% of the articles on PubMed that cite one paper also cite the other.
- *Rate of Adjacent Co-citation* measures the extent to which credit is split. It can also be used as a measure of within-twin similarity. In practice, a rate of adjacent co-citation of 100% means that the two papers are cited in the same parentheses or adjacently at least once in 100% of the follow-on articles that cite both papers, and for which full text was available in open access via PubMed.
- *# of Co-Citing Articles w/ Full Text Observed* corresponds to the number of co-citing articles for which the full text could be retrieved. It indicates the preciseness with which the *Rate of Adjacent Co-citation* could be computed.
- *Back-to-Back (Yes=1)* is another measure of within-twin similarity. It is an indicator variable taking the value of “1” if the two manuscripts were submitted to the same journal and the editor

decided that they were similar enough that they should be published back-to-back in the same issue, and “0” otherwise.

- *Av. Rank Difference* measures semantic similarity. It uses the PubMed keyword-based list of “similar articles” to calculate how far article 1 ranks from article 2, and vice versa. In some cases, it was possible to estimate the rank difference of article 1 from article 2 but not that of article 2 from article 1, either because it was missing or above 30. In those cases, the average is equal to the rank that could be retrieved.
- *Month Difference* can also be used as a measure of within-twin similarity. It corresponds to the number of months separating the publication of the two papers. Values are missing when the publication month for at least one of the articles could not be retrieved from PubMed.