Linking medical device technologies and product markets

Colleen Cunningham[*] and David Hall[†]

May 2025

Abstract

Linking inventions—typically, patents—to specific product markets is challenging but important for many research questions related to innovation. Focusing on the U.S. medical device industry, we apply the Algorithmic Links with Probabilities (ALP) method to generate probabilistic links between patent-based technology classes (USPTO CPC subgroups) and medical device product markets (FDA CFR numbers). The method involves generating keywords from FDA-defined product market descriptions, searching patents for these keywords, and creating technology-to-market and market-to-technology probabilistic linkages. Our mapping has many potential uses, including characterizing the potential markets of nascent technologies and inferring future innovation-related product market competition.

1 Introduction

Innovation involves, first, invention (i.e., the development of a new technology) and, second, commercialization of the technology in product markets (Schumpeter, 1934). Though these steps are conceptually distinguishable, features of the product market likely affect invention choices. For example, choices affecting the rate and direction of invention likely depend on features of a technology's eventual product market application, such as demand and competition. However, in most settings, systematically linking an invention to its potential product market application(s) is not straightforward for the econometrician.

[*]University of Utah, colleen.cunningham@utah.edu
[†]Independent Researcher, dave22hall@gmail.com

A core challenge is that our usual invention data (patents) are not easily linked to products or product markets.[1] Further complicating the matter, new technologies typically have many potential product market applications. For instance, startup firms may experiment across several potential markets and/or pivot across product market applications (Gans et al., 2019). Inferring linkages from realized product market entry, even if possible, likely would draw a very incomplete map of possible linkages.

Prior research has sought to link inventions to markets. These efforts have typically linked patent-based technology classes broadly to industries quite broadly (Kortum and Putnam, 1997; McGahan and Silverman, 2001; Lybbert and Zolas, 2014). Yet, innovation researchers are often interested in understanding the roles that market phenomena such as competition and demand play in invention choices at a more granular level, i.e., *within* an industry.

To that end, we attempt to link patent-based technology classes to product markets for one industry: the U.S. medical device industry. To do so, we apply an algorithmic approach that generates probabilistic links between technology classes and device product markets. The resulting data provide, for example, a means to analyze which product markets an invention will likely end up competing in once commercialized.[2] Using techniques described below, we probabilistically link medical- device–related USPTO technology classes (at the CPC subgroup level) to device-product markets (at the CFR number level).

This paper contributes by applying the Algorithmic Links with Probabilities (ALP) method developed by Lybbert and Zolas (2014) to link technology classes with product markets in the U.S. medical device industry, and, by way of a detailed exposition of the process, provides a map for others seeking to link technology classes and relatively narrow product markets for other industries.

The rest of our paper proceeds as follows. First, we define medical device technology classes and medical device product markets and provide examples. Second, after a brief overview of the ALP methodology developed by Lybbert and Zolas (2014), we detail our adaptation of it to the medical device industry. We then describe our results and how our linkages performs in terms of coverage of markets and technology classes, and provide some validation tests. We conclude with a discussion of how these data can be used to address questions at the intersection of strategy and innovation.

[1]One exception is in the pharmaceutical industry, for which the U.S. Food and Drug Administration (FDA) publishes the Orange Book, which links patents to commercialized drugs. Even then, such linkages are only available for approved drugs, and only once the drug becomes FDA-approved (in other words, post-commercialization).

[2]Or, conversely, what technologies are may emerge to compete in a given market.

2 Medical device technology classes (patents)

To define and categorize medical device inventions, we use USPTO patent data and the USPTO Cooperative Patent Classification (CPC) scheme. We define medical device patents as all granted patents in which the main CPC class falls under the heading A61: "Human Necessities — Medical or Veterinary Science; Hygiene" — excluding sub-classes A61D (veterinary devices) and A61P (pharmaceuticals). The list of included subclasses is in Table 1. We use granted patent data from 1976 until June 2020 (our crosswalk construction date) downloaded via Patentsview.[3]

Table 1: CPC class A61 subclasses in our medical device patent sample

Subclass	**Description**
A61B	Diagnosis; Surgery
A61C	Dentistry
A61F	Filters implantable into blood vessels; Prostheses
A61G	Transport; Accommodations for patients/disables persons; Operating tables or chairs
A61H	Physical therapy apparatus
A61J	Containers specially adapted for medical or pharmaceutical purposes
A61L	Sterilising materials or objects
A61M	Devices for introducing media into/onto the body
A61N	Electrotherapy; Magnetotherapy; Radiation therapy; Ultrasound therapy
A61Q	Cosmetics or similar toilet preparations

The CPC classification system is nested, starting at the most aggregated level, *sections* (e.g., A), within which are *classes* (e.g., A61), within which are *subclasses* (e.g., A61B), within which are *groups* (e.g., A61B 10), within with are *subgroups* (e.g., A61B 10/06), which are the most disaggregated classification. Each patent is assigned to one (or more) CPC *subgroups*. To define technology classes, we used the main assigned CPC subgroup, e.g., A61B 10/06, Biopsy forceps.[4]

3 Medical device product markets (FDA categories)

To define product markets, we use medical device regulatory categories defined by the Code of Federal Regulations (CFRs). The FDA has developed approximately 1,700 regulatory categories of devices, based on their intended use, grouped into 16 medical specialties,[5]

[3] We used the "patent" and the "cpc_current" files.

[4] We collapse subgroups to the 2-digit level in the main version of our matching data. We found this to be the most disaggregated level that still provided predictive power. However, our method easily allows for different aggregations (i.e., from fully disaggregated subgroups up to classes).

[5] https://www.fda.gov/medical-devices/overview-device-regulation/classify-your-medical-device

and defined in Title 21 of the CFR.[6]. These regulatory categories are a slight aggregation of the more narrow FDA product codes used by some prior research (e.g., Chatterji and Fabrizio (2016)). For example, CFR number 870.1875, "Stethoscope," has four associated FDA product codes: Manual Stethoscope, Electronic Stethoscope, Cranial Sound Monitor, and Lung Sound Monitor. Although less commonly used than product codes, CFR categories define "functional categories" of devices that have the same use in or on the body, such as stethoscope, replacement heart valve, or arthroscope (Stern, 2017). Product codes within a CFR category vary in terms of material, method of delivery, and/or product design, but are all used for the same purpose; hence, we think this best aligns with product markets. Helpfully, CFR categories also have detailed descriptions, which are fundamental to our matching process.

As illustrated in Table 3, CFR numbers identify a type of product (e.g., 870.3925, Replacement heart valve), which is neatly situated within a broader functional category (Prosthetic devices), within a medical specialty (Cardiovascular devices). This nested structure is standardized and allows for simple aggregation to (standardized) broader markets. For example, within Title 21 of the CFR, Chapter I, Subchapter H (Medical Devices), Part 870 is reserved for "Cardiovascular Devices." Part 870 contains five subparts for devices grouped by their primary function: B is diagnostic, C is monitoring, D is prosthetic, E is surgical, and F is therapeutic devices. Within each subpart (e.g., Cardiovascular, Dental, Orthopedic, etc.) proximity of regulation number represents device similarity. Table 3 provides examples of regulation numbers for cardiovascular and orthopedic devices. The first three digits (e.g., 870, 888) represent the medical specialty of the device. The first digit after the decimal place indicates the functional use of the device (the Subpart). In Table 3, "Cement Dispenser" and "Cement Mixer" are similar orthopedic surgical devices, indicated by having the same first *two* numbers after the decimal place (4200 and 4210 respectively). In contrast, "Implantable pacemaker pulse generators" and "Replacement heart valves" are both cardiovascular prosthetic devices (as 870.3xxx), but they are not as similar (3610 and 3925 respectively).

4 Algorithmic Links with Probabilities (ALP)

To link medical device technology classes to medical device product markets we use the algorithmic links with probabilities (ALP) method of Lybbert and Zolas (2014), who used the technique to map patent-based technology classes (IPC) *probabilistically* to industry classification schemes (4-digit SITC and ISIC classifications).

[6]The U.S. CFR "is the codification of the general and permanent rules published in the Federal Register by the executive departments and agencies of the Federal Government" Title 21 of the CFR is reserved for the Food and Drug Administration. See https://www.archives.gov/federal-register/cfr/about.html

Table 2: Medical Device Categories by CFR Part

CFR Part	Medical Specialty
868	Anesthesiology Devices
870	Cardiovascular Devices
862	Clinical Chemistry and Clinical Toxicology Devices
872	Dental Devices
874	Ear, Nose, and Throat Devices
876	Gastroenterology-Urology Devices
878	General and Plastic Surgery Devices
880	General Hospital and Personal Use Devices
864	Hematology and Pathology Devices
866	Immunology and Microbiology Devices
882	Neurological Devices
884	Obstetrical and Gynecological Devices
886	Ophthalmic Devices
888	Orthopedic Devices
890	Physical Medicine Devices
892	Radiology Devices

The intuition behind using the ALP approach for our purposes is that, given a technology, there is some probability distribution across product markets for which the technology could be applied. (Conversely, given a market, there is some probability distribution across technologies which may be applied in that market.) Some technological classes might have many potential applications; some might have few. We apply the ALP method to product markets within the medical device industry to generate probabilistic links between medical device technology classes and medical device product markets as defined above.

We have attempted to automate our process as much as possible, and to outline clear decision rules in case where automation was not possible, to make it easy for others to extend and to apply to different contexts.[7] Our approach has three steps, described in detail below: 1) generate keywords for each product market; 2) search for those keywords in a plausibly linked set of patents and aggregate keyword counts to the technology class level to generate frequencies; and 3) create probabilistic linkages between technology classes and product markets, both markets-to-patents and patents-to-markets.[8]

[7]Using an ALP approach similar to ours in other contexts would require a list of technology classes (with descriptive data for technologies in each class) and a list of plausibly linked product markets (with descriptive information for the markets).

[8]For more detail on the general ALP approach, see (Lybbert and Zolas, 2014).

Table 3: Selected example device markets (CFRs) in Cardiovascular and Orthopedics

Part 870, Cardiovascular Devices
Subpart B, Monitoring Devices
870.2100, Cardiovascular blood flowmeter.
870.2850, Extravascular blood pressure transducer.
Subpart E, Prosthetics Devices
870.3610, Implantable pacemaker pulse generator.
870.3925, Replacement heart valve.
Part 888, Orthopedic Devices
Subpart B, Diagnostic Devices
888.1100, Arthroscope.
888.1500, Goniometer.
Subpart E, Surgical Devices
888.4200, Cement dispenser.
888.4210, Cement mixer for clinical use.

4.1 Generating keywords for product markets

The first task in making our technology class-product market crosswalks is generating a succinct set of keywords that accurately describe each medical device product market (of 2,000+). This task is not simple: the keywords must be narrow enough to delineate among product markets, but general enough that searching for market-specific keywords in patent text yields relevant results. To achieve this balance, we trained research assistants (RAs) to manually generate keywords using the CFR product market descriptions. Our RAs were provided with the description text and instructed to identify words specific to the related device type (i.e., those describing the device function), the relevant domains of application (e.g., words for parts of the body, diseases, or conditions the device addresses, etc.), and additional associated terms that help narrow the scope to the specific market. After generating a list of 3–5 keywords for each product market, the RAs manually verified that these keywords matched with patent abstracts via Google patents. Then the RAs iteratively refined candidate keywords based on whether any patents were returned and the apparent relevance of returned patents to the product market. Table 4 provides three example product market descriptions and the relevant set of generated keywords.

4.1.1 Acquiring Synonyms of initial keywords

We expand the initial set of RA-generated keywords by including synonyms from the Unified Medical Language System (UMLS) Metathesaurus. The UMLS, a database of medical

Table 4: Example keywords

CFR Number	Description and RA keywords
§ 868.2375	"A **breathing** (ventilatory) **frequency monitor** is a device intended to measure or **monitor** a patient's respiratory rate. The device may provide an audible or visible alarm when the respiratory rate, averaged over time, is outside operator settable alarm limits." *Terms: "breathing," "monitor," "frequency"*
§ 870.2100	"A **cardiovascular blood flowmeter** is a device connected to a **flow transducer** that energizes the transducer and processes and displays the blood flow signal." *Terms: "blood," "flowmeter," "flow transducer," "cardiovascular"*
§ 872.3500	"**Polyvinylmethylether maleic anhydride** (PVM-MA), **acid copolymer**, and **carboxymethylcellulose sodium** (NACMC) **denture adhesive** is a device composed of **polyvinylmethylether maleic anhydride**, **acid copolymer**, and **carboxymethylcellulose sodium** and intended to be applied to the base of a denture before the denture is inserted in a patient's mouth to improve denture retention and comfort." *Terms: "denture adhesive,", "acid copolymer," "polyvinylmethylether maleic anhydride," "carboxymethylcellulose sodium"*

terminology established in 1986, is maintained by the United States National Library of Medicine. This database is commonly used in medical informatics (see, for example, Bodenreider 2004; Wu et al. 2012; and Carrell et al. 2017). The advantage of using the UMLS over a general purpose thesaurus such as WordNet is its focus on medical terminology. The UMLS brings together over 1 million distinct concepts from 214 incorporated and maintained medical dictionaries across 25 languages, thus providing a comprehensive and detailed set of medical terminology. Burgun and Bodenreider (2001) found that when looking at terminology related to health disorders, the UMLS contained 83% of the terms found in WordNet, whereas WordNet contained only 2% of the terms contained in the UMLS. Table 5 illustrates the usefulness of using the UMLS Metathesaurus to generate keyword synonyms for our purposes. Although synonyms were not found for all terms and markets, we often found multiple useful synonyms for our initial RA-generated medical keywords. The set of RA keywords and UMLS synonyms for each market comprised our final set of keywords.[9]

[9] One challenge in developing the set of keywords was balancing the need for specificity in the terms without producing too few or no results. Initially we tried to incorporate bi-grams to achieve specificity, but when this resulted in too narrow of a search (i.e., no patents were found), we broke out bi-grams into (co-occuring) uni-grams. The first example in Table 2 illustrates as much. Rather than requiring "blood flowmeter" to appear in the text, our search process requires "flowmeter" (what the device is) *and* "blood"

Table 5: Examples of Synonyms Acquired from the UMLS

Product Market	Example RA Keyword	UMLS Synonym
§ 868.2375	"breathing"	"respiratory inspiration" "inhaling" "inhalation"
§ 870.2100	"flowmeter"	No synonyms returned
§ 872.3500	"denture adhesive"	"dental adhesives" "luting agents" "orthodontic adhesives" "dental cements"

4.2 Matching keywords to patents

The next step is searching medical device patents for each set of market-defining keywords. We used the main patent subgroup of each market-matched patent to generate a tabulation of keyword-based matches between product markets and technology classes.

Table 6 includes, for each of our example product markets from our prior tables, one linked technology class to illustrate the matching process. For instance, the keywords for product market "868.2375, Breathing frequency monitor," returned 63 total medical device patents. The most frequent patent subgroup returned was "A61B5/08, Detecting, measuring or recording devices for evaluating the respiratory organs," which comprised 11 patents. In that patent subgroup, 3,472 patents matched to product markets. In Table 6 and the tables that follow, the count of patents listed under technology j that contained the keywords from market i is indicated by m_{ij}, the total count of patents containing the keywords from market i is indicated by Mi, and the total count of patents listed under technology j that matched to any market is indicated by Nj.

Table 6: Example (selected) matching counts

CFR Number	Patent Subgroup	m_{ij}	M_i	N_j
868.2375	A61B5/08	11	63	3,472
870.2100	A61B8/06	11	27	2,878
872.3500	A61K6/30	86	2,258	791

(a term describing the type of flowmeter) to appear, but not immediately together.

4.3 Generating probabilistic links

Using the raw counts outlined in Table 6, we create two directional linkages: (1) a probable technology, given a market; and (2) a probable market, given a technology.

4.3.1 Raw probability weights

Table 7 presents some selected examples to help illuminate how we calculate our raw probability weights, and to illustrate the directional difference. In Panel A, we include the three product markets we have been using as illustrative examples thus far and, for each, the technology class that represent their largest "probable technologies."

The raw weights W_{ij} represent the conditional probability of technology given a market (m_{ij}/Mi). For example, the raw weight W_{ij} linking product market 870.2100, "Cardiovascular blood flowmeter," to patent subgroup A61B8/06, "Measuring blood flow," is 0.407. In other words, their is a 40% probability that a technology in the market "Cardiovascular blood flowmeter" came from the technology class A61B8/06 .

Table 7: Directional Probabilistic Linkages

Panel A: Selected markets

CFR Number	Patent Subgroup	m_{ij}	M_i	N_j	W_{ij}	Adj W_{ij}	Z_{ij}	Adj Z_{ij}
868.2375	A61B5/08	11	63	3472	0.175	0.487	0.003	0.000
870.2100	A61B8/06	11	27	2878	0.407	0.932	0.004	0.134
872.3500	A61K6/30	86	2258	791	0.038	0.203	0.038	0.142

Panel B: Selected patent subgroups

Patent Subgroup	CFR Number	m_{ij}	M_i	N_j	W_{ij}	Adj W_{ij}	Z_{ij}	Adj Z_{ij}
A61B5/08	§ 868.1400	17	39	3472	0.436	1.000	0.005	0.538
A61B8/06	§ 882.1240	246	6918	2878	0.036	0.083	0.085	0.261
A61K6/30	§ 872.3275	86	2193	791	0.039	0.206	0.109	0.147

The raw weights Z_{ij} represent the conditional probability of a market given a technology (m_{ij}/Nj). For example, the raw weight linking Z_{ij} patent subgroup A61B8/06, "Measuring blood flow," to product market 870.2100, "Cardiovascular blood flowmeter," is 0.004. In other words, the probability that a technology in the A61B8/06 subgroup is linked to the "Cardiovascular blood flowmeter" product market is 0.04%.

Just this one example linkage makes it clear that directionality matters when considering probabilities. Specifically, while there is a 40% probability that the technology underlying a "Cardiovascular blood flowmeter" device is from patent subgroup A61B8/06, there is just a

0.04% probability that a technology in the patent subgroup will end up in a "Cardiovascular blood flowmeter" device.

To provide some additional examples, Table 7 Panel B uses the most probable technology classes for each of our example product markets and provides, as example linkages, their most "probable markets" (Z). We can see that, for patent subgroup A61B8/06, the most probable market (Z_{ij} = 0.085 or 8.5%) is 882.1240 "Echnoencephalograph," which is an ultrasonic scanning device that measures blood flow velocity to and in the head.

4.3.2 Adjusting the probability weights

Following Lybbert and Zolas (2014), we adjust these raw probability weights to account for how specifically a technology maps to product markets (for W) — and how specifically a product market maps to technologies (for Z).

Some technology groups (markets) may be inherently more diverse in their mapping to markets (technology groups). To account for this diversity, the adjustment imposes the condition that each market has the same *ex ante* probability of matching with each of the J patent subgroups (and, vice versa, that each patent subgroup has the same *ex ante* probability of matching with each of the I markets). In practice, this increases the weights of technologies (markets) that apply narrowly and decreases the weights of technologies (markets) that apply broadly.[10] As a last fine-tuning adjustment in our matching, to minimize false positives, we set adjusted probability weights below 0.05 to 0 and re-normalize the weights to sum to 1. Table 7 includes, for the most probable matches, the adjusted probability of a technology, given a market (Adj W), and the adjusted probability of a market, given a technology (Adj Z) for each of our example markets and technology classes.

Table 8: Stylized example of adjusted weighting (Table 3 in Lybbert and Zolas)

Market i	Technology j	m_{ij}	M_i	N_j	W_{ij}	Adj W_{ij}	Z_{ij}	Adj Z_{ij}
1	X	98	998	100	0.098	0.540	0.048	1.000
1	Y	900	998	9900	0.902	0.460	0.041	0.083
2	X	2	9002	100	0.000	0.000	0.000	0.000
2	Y	9000	9002	9900	0.999	1.000	0.454	0.917

[10]Specifically, this adjusted weights are calculated using the following:

$$\textit{Adjusted } W_{ij} = \frac{Z_{ij}(W_{ij}/J)}{Z_{i1}(W_{i1}/J) + ... + Z_{iJ}(W_{iJ}/J)}$$

$$\textit{Adjusted } Z_{ij} = \frac{W_{ij}(Z_{ij}/I)}{W_{1j}(Z_{1j}/I) + ... + W_{Ij}(Z_{Ij}/I)}$$

To provide more intuition for the adjustments, Table 8 contains a stylized example building from Lybbert and Zolas (2014). In this simple stylized example, Market 1 had 998 matching patents, the vast majority of which (900 out of 998) were from Technology Y. This results in a relatively larger weight for Technology Y compared to Technology X. However, the adjusted weights take into account that even though Technology X represented a small share of the total matches to Market 1, it matched very narrowly to Market 1; 98% of the Technology X patent were in Market 1. This weighting increases the importance of Technology X (the adjusted weight for Technology X thus increases to 0.540 from 0.098).

5 Results

Continuing with our three example markets, Table 9 lists all technologies meaningfully linked to these markets, i.e., for a given market, the probable technologies. Because these are probabilities, Adj W_{ij} for each market i sums to 1. A relatively higher value of Adj W_{ij} suggests that a particular technology is more probable in that market, relative to the other technologies.

For our three example technology classes, Table 10 lists all markets meaningfully linked to these technologies, i.e., for a given technology, the probable markets. Again, because these are probabilities, Adj Z_{ij} for each technology class i sums to 1. A relatively higher value of Adj Z_{ij} suggests that particular market is more probable for that technology, relative to the other markets.

5.1 Coverage of markets and technologies

Our ALP-based matching approach resulted in a high share of product markets being linked to at least some technologies, but it did not produce complete coverage.

Table 11 provides the total number of product markets within a medical specialty (i.e., 3-digit CFR number), and the share of those markets that were successfully linked to technologies. For example, Anesthesiology has 145 total product markets (i.e., CFR numbers), of which 125 (86%) were successfully assigned probabilistic links to 140 total technologies (i.e., CPC subgroups). However, this calculation assumes the most disaggregated, or most narrow, view of product markets. Because CFR numbers are nested, as described above, these narrow markets may be aggregated up to broader markets, which would increase the coverage considerably.

Table 12 provides the total number of CPC subclasses, and the share of subgroups within each subclass linked to at least one product market. For example, CPC class A61B, which covers technologies for diagnosis and surgery, has a total of 145 subgroups, of which 125 (86%) were linked to at least one of 140 product markets. Again, aggregating up within the

Table 9: Examples of market-technology links

CFR	Adj W_{ij}	Patent Subgroup	Subgroup Description
Breathing frequency monitor			
868.2375	0.487	A61B5/08	Detecting, measuring or recording devices for evaluating the respiratory organs
	0.223	A61B5/72	Signal processing specially adapted for physiological signals or for diagnostic purposes
	0.148	A61M16/00	Devices for influencing the respiratory system of patients by gas treatment
	0.072	A61M16/02	Devices for influencing the respiratory system of patients by gas treatment, electrical means
	0.070	A61B5/11	Measuring movement of the entire body or parts thereof
Cardiovascular blood flowmeter			
870.2100	0.932	A61B8/06	Measuring blood flow
	0.068	A61B8/02	Measuring pulse or heart rate
Denture adhesive			
872.3500	0.547	A61K6/88	Preparations for artificial teeth, for filling teeth, or for capping teeth
	0.250	A61Q11/00	Preparations for care of the teeth, of the oral cavity, or of dentures
	0.203	A61K6/30	Compositions for temporarily or permanently fixing teeth or palates, e.g., primers for dental adhesives

Table 10: Examples of technology–market links

Patent Subgroup	Adj Z_{ij}	CFR	CFR Device Description
Detecting, measuring, or recording devices for evaluating the respiratory organs			
A61B5/08	0.538	868.1400	Carbon dioxide gas analyzer
	0.462	868.5690	Incentive spirometer
Measuring blood flow			
A61B8/06	0.261	882.1240	Electroencephalograph
	0.215	882.1925	Ultrasonic scanner calibration test block
	0.215	890.5300	Ultrasonic diathermy
	0.176	870.2880	Ultrasonic transducer
	0.134	870.2100	Cardiovascular blood flowmeter
Preparations for artificial teeth, for filling teeth, or for capping teeth			
A61K6/30	0.147	872.3275	Dental cement
	0.142	872.3410	Ethylene oxide homopolymer and/or CMC sodium denture adhesive
	0.142	872.3420	CMC sodium and cationic polyacrylamide denture adhesive
	0.142	872.3480	Polyacrylamide polymer (modified cationic) denture adhesive
	0.142	872.3490	CMC sodium and/or PVM calcium-sodium double salt denture adhesive
	0.142	872.3500	PVM-MA, acid copolymer, and NACMC denture adhesive
	0.141	872.3450	Ethylene oxide homopolymer and/or karaya denture adhesive

Table 11: Market coverage by medical specialty

Part: Specialty	Total Markets	Markets Linked	Coverage	Technologies
868: Anesthesiology	145	125	0.86	140
870: Cardiovascular	151	128	0.85	129
862: Clinical Chemistry/Toxicology	243	123	0.51	177
872: Dental	135	114	0.84	178
874: Ear, Nose, And Throat	62	42	0.68	58
876: Gastroenterology-Urology	82	70	0.85	126
878: General And Plastic Surgery	94	84	0.89	188
880: General Hospital And Personal Use	113	88	0.78	186
864: Hematology And Pathology	116	79	0.68	146
866: Immunology And Microbiology	214	86	0.40	100
882: Neurological	125	105	0.84	160
884: Obstetrical And Gynecological	101	80	0.79	127
886: Ophthalmic	135	109	0.81	108
888: Orthopedic	90	82	0.91	74
890: Physical Medicine	85	72	0.85	111
892: Radiology	83	60	0.72	80

nested structure of technology classes would increase coverage.

5.2 Validation

We have undertaken to generate a mapping between technology classes and the medical device markets. In an ideal world, we would have some means to evaluate the extent to which our data match "reality," that is, to determine both the false positives and the false negatives.

Determining each type of error presents its own challenge. To identify false positive links between patents classes and CFR numbers in our data (i.e., a link that our algorithm indicates is probable but is not in reality a link), one would need a source of the universe of *realized* and *unrealized* applications of a patent to a medical device market. In other words, a single invention might have multiple possible applications. However, any data we might obtain would only have the market to which the technology was actually applied. A data source containing *all* possible outcomes, as ours attempt to do, does not exist, as far as we are aware. If it did, we would not need to create this cross walk. This issue makes identifying false positives an impossible task.

Identifying false negatives—probable links between technology classes and markets that do not appear in our data—is more practical in theory but has its own challenges in practice.

Table 12: Technology coverage by main CPC subclass

Main CPC	Total sub-groups	Subgroups Linked	Coverage	Markets
A61B Diagnosis; Surgery	283	200	0.71	599
A61C Dentistry	125	78	0.62	96
A61F Filters implantable into blood vessels; Prostheses	156	115	0.74	342
A61G Transport, Accommodations for patients/disables persons, Operating tables or chairs	117	42	0.36	57
A61H Physical Therapy Apparatus	87	36	0.41	74
A61J Containers for medical or pharmaceutical purposes	59	23	0.39	32
A61K Preparations for medical, dental, or toilet purposes	588	302	0.51	392
A61L Sterilising materials or objects in general	228	96	0.42	154
A61M Devices for introducing media into/onto the body	189	77	0.41	282
A61N Electrotherapy; Magnetotherapy; Radiation therapy; Ultrasound therapy	35	25	0.71	114
A61Q Cosmetics or similar toilet preparations	37	11	0.30	24

In theory, one should be able to take observed outcomes of patents that were applied to markets and then compare them to the ALP links to see what links, if any, we are missing. Yet, these observed outcomes are hard to find. We identified two potential external sources of data that might provide such observed outcomes: so-called virtual patent markings, and CFR mentions in patent text. Each of these potential sources introduced challenges to validating our data, due to the quality and/or precision of the observed matches in these external sources.

5.2.1 Patent Markings

Our first approach to generating observed patent-product links is to leverage "virtual patent markings" provided by individual firms. Some medical device firms provide public lists of at least some of their products with the patents intended to protect those products, i.e., the

Table 13: Examples of Seemingly Distant Patent–Product Links in Patent-Marking Data

Patent Group	Description	CFR Market	Description
A61B17	Surgical instruments, devices or methods, e.g., tourniquets	§ 888.3060	Spinal intervertebral body fixation orthosis
A61B17	Surgical instruments, devices or methods, e.g., tourniquets	§ 888.3520	Knee joint femorotibial metal/polymer non-constrained cemented prosthesis
A61L11	Methods or apparatus for sterilising materials or objects in general; specially adapted for refuse	§ 878.4780	Powered suction pump

patent markings.[11] We assembled a dataset of patent markings for six firms: Medtronic, Boston Scientific, Zimmer Biomet, Stryker, Fresenius, and Philips NV. Doing so required a combination of scraping websites and extracting text from PDF files. We then used fuzzy matching[12] to match FDA-approved device names to the product names listed by the firms in their patent markings. This process resulted in 65 unique FDA approved devices linked to 221 unique patents. Notably, this sample is small and represents only 8 of the 16 medical specialties in our data.

Overall, our crosswalk predicted 51% of the links in the marking data. In inspecting the links missing in our crosswalk, we found many of the links were not obvious matches from available data. This disconnect results, in part, because the patent markings lists include fairly complex devices, with many patents associated with them in the firm's patent marking files. It is possible that these additional patents cover a narrow portion of the product, thereby increasing the distance between the main patent classification of the associated patent(s) and the relevant market of the device. Table 13 provides examples of some of these seemingly distant matches that do not appear in our data but do appear in the patent-marking data. For example, patent group A61B17, "Surgical instruments, devices or methods, e.g., tourniquets," was found to be linked to CFR "888.3060, Spinal intervertebral body fixation orthosis."

[11] For example, the patent markings for Boston Scientific can be found at https://www.bostonscientific.com/en-US/patents.html.

[12] We used the Jaro-Winkler distance from the stringdist package in R, limiting matches to those matches with a distance below 0.1.

5.2.2 Patents Listing CFR Numbers

Second, we explored the possibility that medical device inventors explicitly name relevant CFR numbers in the patent. To investigate this idea, we searched for the phrase "code of federal regulations" on Google Patents and then searched for medical-device CFR numbers in the full patent text. Two issues arose in trying to identify and extract this information. First, searching the full patent text for every medical device CFR was computationally expensive. We therefore started by sampling.[13] We found in our initial samples that very few patents contain medical device CFR numbers or reference to CFR. Second, the few patents that mention CFR numbers imprecisely list applicable markets. For example, patent application US20040138688A1[14] lists all CFR numbers under CFR Part 862 (Clinical Chemistry/Toxicology), a list of over 220 CFR numbers. This list is prefaced with the statement, "The embodiments disclosed herein can be employed in conducting *at least some* of the tests enumerated below" (emphasis added). This imprecision makes sense if inventors want to seek as broad of coverage as possible, and referencing a broader range of market applications can help. Given the rareness and imprecision of in-patent CFR mentions, we decided this was not a fruitful path for constructing a validation dataset.

6 Discussion and Conclusion

Research at the intersection of strategy and innovation can benefit from data that links inventions to product markets. However, the link between patents and markets is elusive in most cases because first, firms typically do not disclose the patents used in a given product, and second, because the relationship between technologies and products is not one-to-one. To overcome these challenges, we developed an application of the Algorithmic Links with Probabilities (ALP) approach to link technology classes to product markets in the U.S. medical device industry.

We believe such a linkage can be useful for a variety of future questions related to strategy and innovation. By providing a many-to-many mapping of technologies and markets, we can provide insight into several questions, for example, what are the relevant potential markets for a startup firm with a nascent invention? Further, even for firms who succeed a commercializing an invention into a particular product market, our data open a window into which other markets they might have reasonably entered but did not. Additionally, our ALP approach provides a way to infer the breadth of firms' technological capabilities, not just in a technical sense (i.e., the breadth of technology classes in their portfolio), but in terms of

[13] 4,200 resulting patents X 2,000 CFR numbers results in over 8 million searches. Assuming 0.25 seconds per search, this would be a month of non-stop compute time to search (assuming no way to parallelize the process).

[14] https://patents.google.com/patent/US20040138688A1/en

the product market (i.e., the breadth of potential markets that firms can enter given their existing technological capabilities). These are just a few examples of how such a linkage may be useful for studying innovation.

We hope other researchers find our efforts useful. We also hope this paper has provided a map for other researchers to apply the ALP methodology to other industries and contexts.

References

Bodenreider, Olivier, "The unified medical language system (UMLS): integrating biomedical terminology," *Nucleic acids research*, 2004, *32* (suppl_1), D267–D270.

Burgun, Anita and Olivier Bodenreider, "Comparing terms, concepts and semantic classes in WordNet and the Unified Medical Language System," in "Proceedings of the NAACL'2001 Workshop,"WordNet and Other Lexical Resources: Applications, Extensions and Customizations" 2001, pp. 77–82.

Carrell, David S, Robert E Schoen, Daniel A Leffler, Michele Morris, Sherri Rose, Andrew Baer, Seth D Crockett, Rebecca A Gourevitch, Katie M Dean, and Ateev Mehrotra, "Challenges in adapting existing clinical natural language processing systems to multiple, diverse health care settings," *Journal of the American Medical Informatics Association*, 2017, *24* (5), 986–991.

Chatterji, Aaron K and Kira R Fabrizio, "Does the market for ideas influence the rate and direction of innovative activity? Evidence from the medical device industry," *Strategic management journal*, 2016, *37* (3), 447–465.

Gans, Joshua S, Scott Stern, and Jane Wu, "Foundations of entrepreneurial strategy," *Strategic Management Journal*, 2019, *40* (5), 736–756.

Kortum, Samuel and Jonathan Putnam, "Assigning patents to industries: tests of the Yale technology concordance," *Economic Systems Research*, 1997, *9* (2), 161–176.

Lybbert, Travis J and Nikolas J Zolas, "Getting patents and economic data to speak to each other: An 'algorithmic links with probabilities' approach for joint analyses of patenting and economic activity," *Research Policy*, 2014, *43* (3), 530–542.

McGahan, Anita M and Brian S Silverman, "How does innovative activity change as industries mature?," *International Journal of Industrial Organization*, 2001, *19* (7), 1141–1160.

Schumpeter, Joseph Alois, *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*, Transaction Publishers, 1934.

Stern, Ariel Dora, "Innovation under regulatory uncertainty: Evidence from medical technology," *Journal of public economics*, 2017, *145*, 181–200.

Wu, Stephen T, Hongfang Liu, Dingcheng Li, Cui Tao, Mark A Musen, Christopher G Chute, and Nigam H Shah, "Unified Medical Language System term occurrences in clinical notes: a large-scale corpus analysis," *Journal of the American Medical Informatics Association*, 2012, *19* (e1), e149–e156.