

II. Introduction

CHI Research, Inc. was engaged by the National Science Foundation through a subcontract to Abt Associates to undertake a bibliometric analysis of the emerging area of tissue engineering with the intent of describing quantitatively NSF's role in the area and examining the co-authorship structure of the field.

The project consisted of five parts:

1. Identifying core papers fundamental to tissue engineering
2. Constructing a database of information on the papers
3. Analyzing the nature and extent of NSF's overall role in the field as revealed through funding acknowledgements on papers
4. Developing representations of coauthorship information for leading authors in the field, with indications of NSF's presence.
5. An analysis of international patenting in tissue engineering

This report describes each of these steps in turn. The first section describes how the foundation for the analysis was carefully laid through development of a sophisticated methodology developed specifically to identify core papers fundamental to tissue engineering. After this, a basic description is offered of the growth of tissue engineering, as revealed in the full paper set and in a special set of papers that use the term "tissue engineering" in their abstract or title. There follows a quantitative analysis of NSF's role as revealed through acknowledgments of funding reported on papers. Then maps and tables are presented that together reveal the patterns of coauthorship in the field and NSF's presence within the oeuvre of leading authors. Finally an analysis of international patenting in tissue engineering is reported.

III. Methodology

A. Finding core papers fundamental to tissue engineering

The fundamental methodological work in this project was to devise a way of identifying core papers fundamental to tissue engineering. This was very challenging. CHI needed to identify the papers in a rapidly evolving area that brings together a heterogeneous set of technologies and research approaches, and in which no two researchers seem to agree on a definition. At some level, all biomedical knowledge not directly concerning disease probably will contribute to tissue engineering. However, time was limited, so every biomedical paper could not be assessed for relevance to tissue engineering.

The approach we adopted was to develop a "gold standard" method to find papers. The method had two stages. In the first stage, we identified manually papers and patents in a narrowly defined core area, which could be nothing else, but tissue engineering and that all would agree were tissue engineering. We defined the core of tissue engineering to be seeding autologous cells and growth factors onto three-dimensional biodegradable scaffolds with the aim of forming new functional tissue. We found core papers by developing filters (combinations of keywords and classifications used to search PubMed or the USPTO databases, see Appendices 1 & 2) to identify papers and patents that met the definition, then reading abstracts and titles to screen documents found by the filters.

In the second stage, bibliometric links were used to find papers seen by at least two researchers as relevant to the core. This was implemented by finding papers cited in the core patents by at least two inventors or in the core papers that were reviews of tissue engineering by two authors. Complications were added to the citation element of the method by the need to consider coauthorship. For if one group habitually cites a paper, should that count as a paper that two tissue engineering authors agree is tissue engineering? We implemented a strict version of the criterion in which at least two groups had to cite a paper. We did this by choosing papers whose number of citations exceeded the number of citations from the most citing inventor/author. This is not perfect, and one could imagine situations in which it fails. However, a search for the perfect criterion quickly gets extremely complex and expensive. Note that our criterion works to give a lot of power to those who wrote just one review of the field. Their "votes", as expressed in papers they cited in their review, carry as much weight as 5 or 10 "votes" from papers of prolific review writers.

Using this method, 1,824 core papers fundamental to tissue engineering were identified. Table 1 describes how these 1,824 papers were obtained.

Table 1 – The construction of the base set of core papers fundamental to tissue engineering

Number of candidate papers	How they were found	Number of TE papers	How they were chosen
1,814	filtered from PubMed ¹	872	found to be TE upon reading abstracts and titles
	165 of these are review papers		
5,051	cited in review papers	783	of these are cited by at least 2 authors ²
2,009	cited in TE patents (266)	221	of these are cited by more than 2 inventors ³
		330	additional papers were cited in both a patent and paper
		1,824	after duplicates are removed.

We designed the filter and associated paper gathering strategy to address the challenges inherent in defining tissue engineering. In an area where no two scientists completely agree on the boundaries, CHI's methods require that we begin work with an explicit definition, which we make public. No doubt, scientists in the area who do not agree with each other on a definition can all agree with each other that they disagree with the CHI definition. Nevertheless, as we read the abstracts of papers, we were pleased to find that our definition was in line with definitions found in abstracts whose authors made statements of the kind: "tissue engineering is . . .".

Note also the importance of the citation component. Citations from review papers and from filtered patents were used to find papers. This element makes the judgments of the scientific community central to the decision to include a paper or not. 66% of the papers included in the study entered because of citation links. Only 33% entered solely through the paper filter.

In some sense, most of biomedical knowledge except disease diagnosis and treatment can be related to tissue engineering. Tissue engineering builds most directly on: cryopreservation,

¹ Review and research papers only

² Excluding coauthorship, i.e. the number of citations to the papers overall exceeded the number from the most citing author.

³ Excluding coinvention, i.e. the number of citations to the papers overall exceeded the number from the most citing inventor.

development of bioreactors, cell culture techniques, understanding of growth factors, peptides, collagen, fibrin, polymers, development of biomaterials, understanding of cell growth and differentiation, knowledge of how nerves, blood vessels, bone, heart, bladder, liver and skin all work - and no doubt more besides. All of this should really be included to capture all the knowledge that goes into TE. However, an unfocused study of all of biomedicine except disease diagnosis and treatment would not be practical or useful.

So we focus. At its heart, the filter focuses on those who brought things together, combined elements, in the ways they needed to be combined to do tissue engineering. It highlights synthesis work. Thus, those who may have had the vision of such a synthesis earlier and may have pushed it harder might appear more prominently.

Synthesizing elements means bringing together disparate high-level expertise. So perhaps the emphasis on synthesis is related to the highly collaborative nature of the work included. We might also expect that the most ardent synthesizers would be the most highly collaborative.

The emphasis on synthesis will create the appearance of incompleteness from an individual scientist's perspective. This is because to participate an individual has to have a highly relevant skill and knowledge set, for example, cardiovascular fluid dynamics. Understanding cardiovascular fluid dynamics is crucial to building blood vessels. But it is really the point at which expertise in cardiovascular fluid dynamics is applied to building new vessels that we are trying to capture. To build new blood vessels will require more than even being the world expert on cardiovascular fluid dynamics - hence the idea that synthesis of this expertise with something else is crucial to the work that got pulled into the paper set here.

Synthesis work is clearly much harder to identify than say, all work on liver or on growth hormone X. So although CHI always goes to great lengths to search properly and go for 95% of what is out there, here it may well not have been possible to achieve that high a percentage. Because citations were used to find many of the papers, we expect that more cited papers are more likely to be included and uncited papers less likely.

B. Constructing the analysis paper set

There are 1,824 papers in the tissue engineering set, and 1,056 in the analysis set. The construction of the analysis set and the reason for the difference in size are described in this section.

The first reason that the analysis set is smaller is that we are analyzing only US-authored papers and the full 1,824 tissue engineering set contains both foreign and domestic papers. The second reason the analysis set is smaller is that it contains only papers for which we obtained full information. We had bibliographic references only for each paper in the tissue engineering set. To conduct our analysis we needed to obtain for each US-authored paper a complete set of information including: all authors, all institutions, and all funding sources acknowledged on the paper.

CHI combined several sources to construct this information. For most papers, complete author and institution information was bought from ISI.⁴ However, NSF was concerned not to limit the study to papers on which ISI could provide information, therefore it was necessary to look up institutional information for some papers in the library. Funding information was obtained from CHI's database of funding acknowledgements on US-authored papers cited in patents. However, library work was needed to obtain funding information on quite a few papers that were not already in this database.

Inevitably, all three pieces of information could not be obtained for every paper and the size of the set was reduced further. Papers were only looked up if there were at least two papers in the same journal. This necessary economy eliminated most papers in obscure journals that were not in libraries. Nevertheless, some papers could not be found because the journal was not available, or because the volume was missing or because the reference was so incorrect that the research assistants had no luck searching for the paper. After the lookup was completed, papers that lacked one or more pieces of information (authors, institutions, funding) were eliminated. Note that if a paper was examined and lacked funding acknowledgments, the paper was not eliminated, rather it was kept and marked as "no funding acknowledged".

The analysis set contains US-authored papers for which we have looked up funding acknowledgments, for which we know who all the authors were, and for which we know all their associated institutions. There are 1,056 papers in the analysis set.

⁴ Note that PubMed does not provide complete institutional information, which is why ISI was used.