



Praxis im Umgang mit medizinischen Forschungsdaten

DCT-Befragungsergebnisse

Lars Müller

Christoph Szepanski

Thomas Wetzel

Hans-Christoph Hobohm

Judith Pfeffing

Fachhochschule Potsdam

Fachbereich Informationswissenschaften

Projekt: DataCreativity Tools for Innovation and Research (DCT)

Förderprogramm: FHProfUnt

Datum: 18.3.2013

GEFÖRDERT VOM



**Bundesministerium
für Bildung
und Forschung**

Der Bericht steht unter folgender Creative-Common-Lizenz:

„Attribution-NonCommercial-ShareAlike 3.0 Unported“ <http://creativecommons.org/licenses/by-nc-sa/3.0/>



Inhaltsverzeichnis

1	Einleitung.....	3
1.1	Fragestellung und Operationalisierung	3
1.2	Durchführung der Befragung.....	4
2	Ergebnisse	5
2.1	Anlässe zur Datenanalyse	5
2.2	Auswahl der Datenquellen	6
2.3	Nutzung externer Hilfsquellen	7
2.4	Darstellung von Daten	8
2.5	Arbeitssituation	9
2.6	Verwendung der Analyseergebnisse	10
2.7	Professioneller Hintergrund der Befragten.....	11
3	Zusammenfassung und Diskussion.....	12
3.1	Ergebnisse	12
3.2	Interpretation.....	12
3.3	Schlussfolgerungen für DCT-Entwicklung.....	12
3.4	Informationswissenschaftliche Diskussion der Ergebnisse	13
3.5	Fazit	15
4	Anhang.....	16
4.1	Interviewleitfaden	16
4.2	Online-Fragebogen	18

1 Einleitung

Das FuE-Projekt *Data Creativity Tools for Innovation and Research* (DCT) hat zum Ziel, aus großen, heterogenen medizinischen Forschungsdatenbeständen die Entwicklung neuer Forschungsfragen und –hypothesen anzuregen. In Kooperation mit dem Projektpartner *Open European Nephrology Science Center* (OpEN.SC) von der Charité Berlin wird hierfür eine Webanwendung entwickelt. Diese Entwicklung zielt auf eine neuartige Informationsumgebung ab, die kreative Datenexploration begünstigen soll. Um die Anwendung nutzerorientiert zu gestalten, wurden qualitative Interviews und eine Onlinebefragung zur Praxis im Umgang mit medizinischen Forschungsdaten durchgeführt.

1.1 Fragestellung und Operationalisierung

Ziel der Onlinebefragung war es, mehr über den derzeitigen praktischen Umgang von Mediziner*innen mit „ihren“ Daten zu erfahren, um zielgruppenangepasste Funktionen für Data Creativity Tools (DCT) entwickeln zu können.

Folgende Fragestellung lag dem Onlinefragebogen zugrunde:

An welchen Stellen im Prozess der Problemfindung und -benennung (Datenexploration, Anomalieerkennung, Hypothesenformulierung) können veränderte Formen der Datenpräsentation, -integration und -vernetzung forschenden Patholog*innen helfen, bei der Formulierung von neuen Forschungsfragen und Hypothesen ihr kreatives Potential besser auszuschöpfen?

Der Operationalisierung lag ein datenorientierter Forschungsprozess, an dessen Beginn die Datenanalyse und an dessen Ende ein neues Forschungsprojekt steht als Prämisse zugrunde. Abweichungen und Übereinstimmungen von diesem Modell zur Praxis sollen Anregungen für Anknüpfungs- und Interventionspunkte beim Systemdesign liefern.

Insgesamt wurden sieben Themenfelder abgefragt:

1. Anlass zur Datenanalyse zur Ermittlung der Ausprägung explorativer Datenanalysepraxis,
2. Wahl der Datenquellen und
3. Anwendung von Datenanalysen zur Ermittlung der Datenorientierung,

4. Verwendung externer Hilfsquellen zur Ermittlung der Wissensumgebung bei der Datenanalyse,
5. Bevorzugte Datendarstellung und
6. Arbeitssituation für Hinweise auf Funktionsanforderungen sowie
7. Professioneller Erfahrungshintergrund der Befragten zur Bewertung der Antworten.

Es wurden Fragen zum praktischen Umgang mit Daten gestellt. Ergebnisse aus den explorativen Interviews, die einen breiten Fokus auf das Informationsgeschehen im medizinischen Arbeitsalltag hatten, sind in die Ausarbeitung des Fragebogens eingeflossen.

1.2 Durchführung der Befragung

Die leitfadengestützten Interviews¹ wurden in der Zeit vom 13.7.2011 bis zum 25.8.2011 geführt. Es wurden insgesamt fünf Mediziner in ca. halbstündigen Einzelgesprächen befragt. Die Antworten sind hier insofern von Belang, als dass sie bei Bedarf zur Bewertung und Anreicherung der Antworten aus der Onlinebefragung herangezogen wurden.

Der Online-Fragebogen wurde über den Webservice <http://www.unipark.info> erstellt und eingesetzt.² Über einen Link war er für die Befragten zu erreichen. Diese wurden über einen internen Mailverteiler der nephrologischen Abteilung der Charité-Berlin angesprochen und um Teilnahme an der Befragung gebeten. Die Antworten wurden im Zeitraum von Mitte Juni bis Mitte September 2012 erfasst.

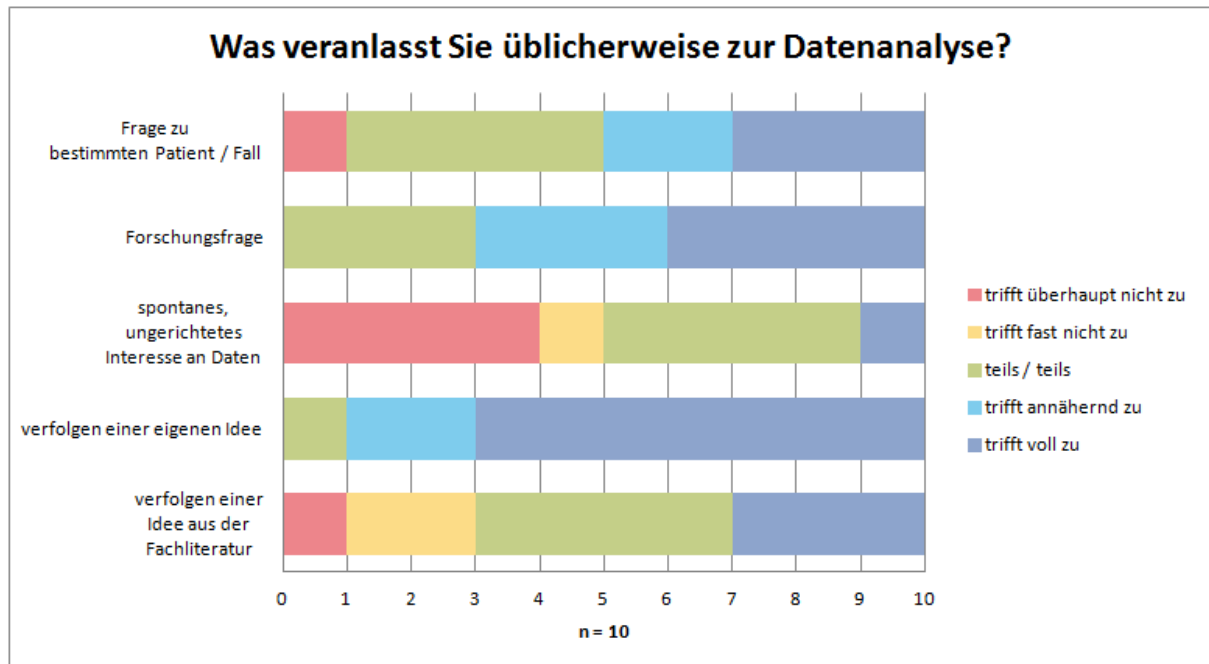
Mit genau 10 vollständig ausgefüllten Fragebögen blieb der Rücklauf deutlich hinter den Erwartungen zurück. Da wir aufgrund der automatisch erfassten Webstatistik mit hoher Wahrscheinlichkeit annehmen können, dass die Antworten aus dem Kern unserer Zielgruppe stammen, lassen sie trotz des geringen Rücklaufs Schlussfolgerungen zu, die wir in die Entwicklung der Data Creativity Tools einfließen lassen. Fraglos ist aber auch, dass die Ergebnisse explorativen Charakter haben. Erst eine Wiederholung der Befragung mit deutlich höherem Rücklauf könnte zuverlässige allgemeine Aussagen über die Datenpraxis von Medizinern gestatten.

¹ Interviewleitfaden im Anhang, S. 16f.

² Fragebogen im Anhang, S. 18ff.

2 Ergebnisse³

2.1 Anlässe zur Datenanalyse

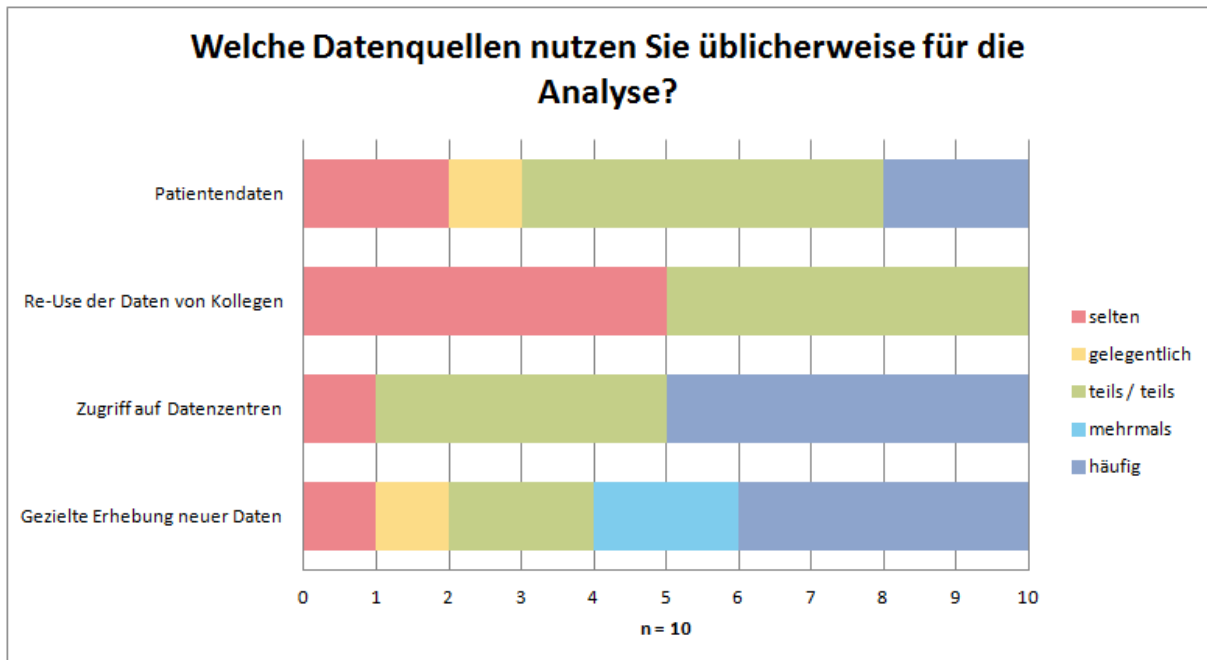


Für die Befragungsteilnehmer überwiegt eine eigene Idee als Anlass zur Datenanalyse. Auch Forschungsfragen, die etwas elaboriertere Versionen von Ideen, geben entsprechenden Anlass. Anregungen aus der Fachliteratur spielen eine geringere Rolle. Die Exploration von Daten aus ungerichtetem Interesse heraus bildet die Ausnahme. Ideen verfolgen wollen, ist Anlass zur Datenanalyse, Ideen finden nicht. Somit zeigen die Antworten, dass Datenanalyse eher hypothesengeleitet als explorativ ist.

Unsere Prämisse besagt, dass für datenorientierte Forschung die Idee aus dem „spontanen, ungerichtetem Interesse“ an den Daten, also der Exploration entspringen muss. Es müsste daher gelingen durch die DCT hier einen „Kurzschluss“ zu erzeugen, bei dem Ideen aus der Datenexploration generiert werden, die für sich Anlass bieten, die Daten gezielter – hypothesengeleitet – zu analysieren.

³ Aufgrund von n=10 werden in den Diagrammen alle Antworten nicht wie üblich in Prozentanteilen, sondern in absoluten Zahlen dargestellt.

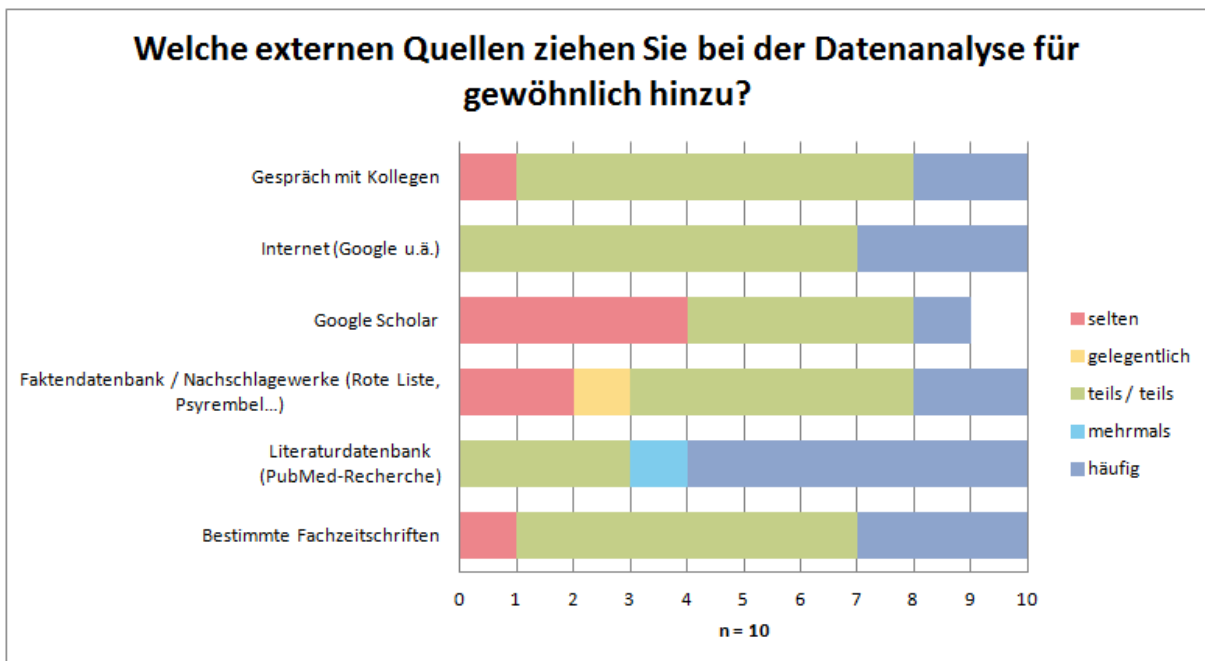
2.2 Auswahl der Datenquellen



Bei der Frage nach der Datenquellenauswahl fällt auf, dass der Re-Use von Daten von Kollegen keinmal als häufigste Option gewählt wurde. Überraschenderweise gibt die Hälfte der Befragten aber Datenzentren als häufig genutzte Quelle an, was ja auch eine Form von Re-Use ist. Die Antworten in Bezug auf Datenzentren beziehen sich vermutlich auch auf die innerhalb der Charité zur Verfügung stehenden verschiedenen Datenzentren, die Teilelemente des Krankenhausinformationssystems (KIS) bilden. Hierzu zählen beispielsweise Teile des krankenhausinternen SAP Systems, das Charité Intranet oder die vom Institut für Laboratoriumsmedizin bereitgestellte Parameter- und Referenzwertdatenbank. Die unterschiedlichen Datenquellen lassen an dieser Stelle keine klare Abgrenzung zum Re-Use von Daten von Kollegen zu. Wir vermuten jedoch, dass der Vorteil in Datenzentren darin gesehen wird, dass dort Daten in der Regel einheitlich aufbereitet und dokumentiert sind sowie (datenschutz-)rechtliche Fragen nach Zugriff und Nutzbarkeit nach festgelegten Verfahren geregelt werden. Gezielte Erhebung neuer Daten steht in der Nennung der Häufigkeit noch vor der Nutzung von Patientendaten. In den qualitativen Interviews hat sich gezeigt, dass es insbesondere einen Bedarf an leichtem, sicheren und nachvollziehbarem Zugang zu Forschungsdaten gibt.

Dies wird hier nochmals bestätigt und lässt vermuten, dass insbesondere in der Aufbereitung von Patientendaten in einem Datencenter, wie es OpEN.SC leistet, Potential liegt. Qualitativ liegt das Potential insbesondere in der verbesserten Integration von Daten aus unterschiedlichen Quellen (andere Standorte, andere Teildisziplinen).

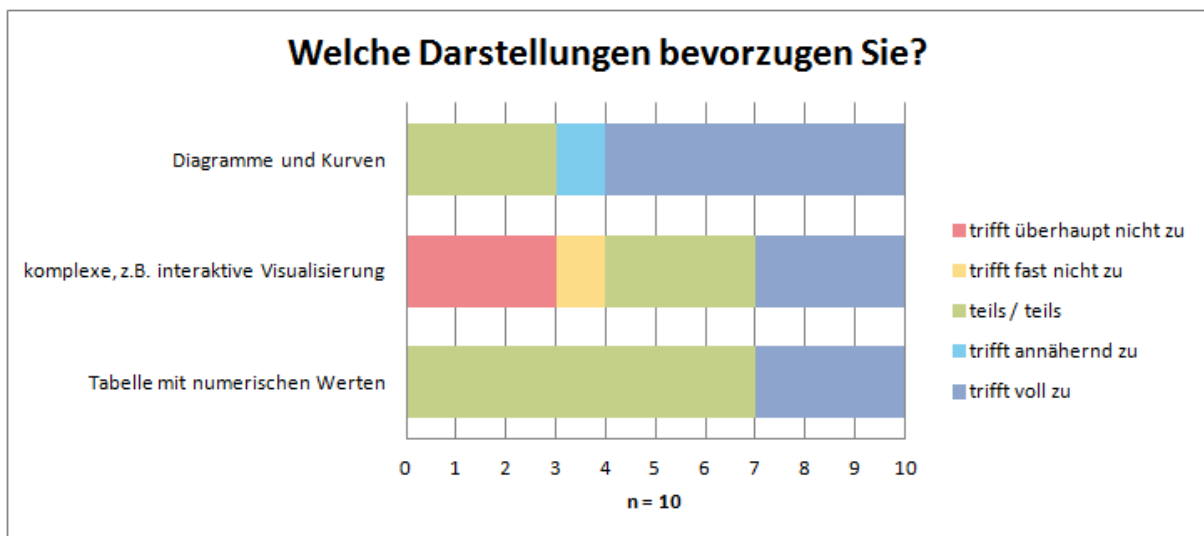
2.3 Nutzung externer Hilfsquellen



Alle Formen externer Informationsquellen sind wichtig, es dominieren klar wissenschaftliche Informationsmedien, PubMed und bestimmte Fachzeitschriften zusammen mit nicht näher differenzierter Internetnutzung. Mit den Nachweisinstrumenten zu Fachpublikationen kann das wissenschaftliche Umfeld zu bestimmten Fragen erkundet werden. Auch Google-Scholar liefert hierfür Links auf Dokumente, wird aber zurückhaltender eingesetzt. Nur ein Befragter gab an, diesen Dienst häufig zu nutzen. Die Intensität der Nutzung fachlicher Nachschlageweisensysteme ist breit gestreut. Welcher Zweck bei der allgemeinen Internetnutzung die von allen Befragten „teils/teils“ oder „häufig“ genannt wurde, verfolgt wird, ist nicht erhoben worden. Antworten aus den Interviews legen jedoch den Schluss nahe, dass hier insbesondere nach aktuellen Entwicklungen gesucht wird. Das persönliche Gespräch rangiert im Mittelfeld.

DCT sollten entsprechend möglichst viele wichtige sekundäre Informationsquellen integrieren können. Für die Erweiterung der Wissensbasis zu bestimmten Ideen ist es hilfreich, auch weniger genutzte Instrumente einzubinden und aktiv anzubieten. Es ergibt sich eine Reihung von a) inhaltlicher Tiefe, b) Aktualität und c) allgemeinem Kontext. Der persönliche Austausch bei der Datenanalyse scheint eine unausgeschöpfte Ressource zu sein.

2.4 Darstellung von Daten

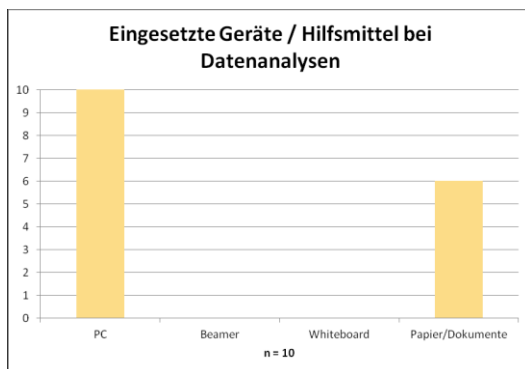
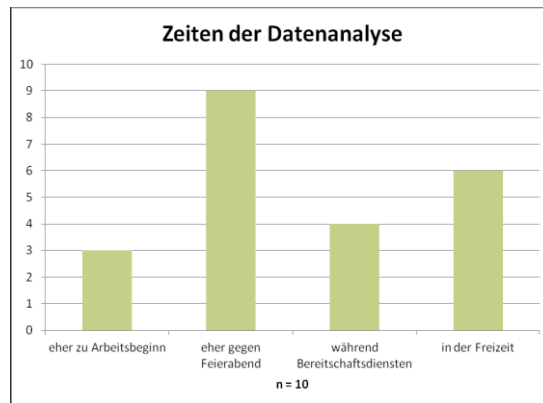
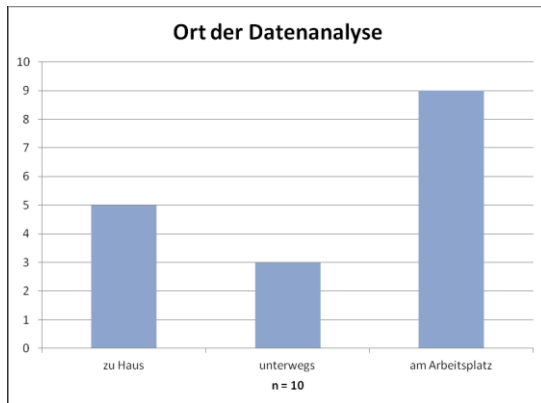


Traditionelle Visualisierungen in Form von Diagrammen und Kurven werden eindeutig von den meisten Befragten bevorzugt. Eine Tabelle mit numerischen Werten wird von allen Befragten zumindest noch teils/teils bevorzugt. Die Haltung gegenüber komplexen Visualisierung ist dagegen sehr gespalten: vier Befragte antworten hier mit „trifft überhaupt nicht zu“ oder „trifft fast nicht zu“. In drei Fällen gehört diese Form aber zu den Bevorzugten („trifft voll zu“). Anspruchsvolle Visualisierungen erhöhen die Komplexität der Darstellung. Dies ist für komplexe Fragestellungen sicher nötig und gewollt. Zugleich ergaben die Interviews aber auch, dass wenig bis keine Zeit und Bereitschaft zur Verfügung steht, sich in die Anwendung von neuen Tools (neuer Software) einzuarbeiten.

Dennoch liegt genau hier Potential für die DCT-Entwicklung. Numerische Daten in Tabellen oder zwei- bis dreidimensionalen Grafiken können die Komplexität der Daten und ihrer Beziehungen nicht immer angemessen abbilden. Insbesondere lassen sich in den traditionellen Grafiken und tabellarischen Darstellung nicht-numerische Werte kaum oder gar nicht visualisieren.

Hier stecken Möglichkeiten in neuen Visualisierungstechniken, deren Wert möglicherweise noch nicht erkannt wurde. Da sie nur in drei Fällen als die bevorzugte Form der Datendarstellung genannt wurde, müssen sie allerdings behutsam, d.h. optional eingesetzt und idealerweise akzeptanzfördernd eingebunden werden.

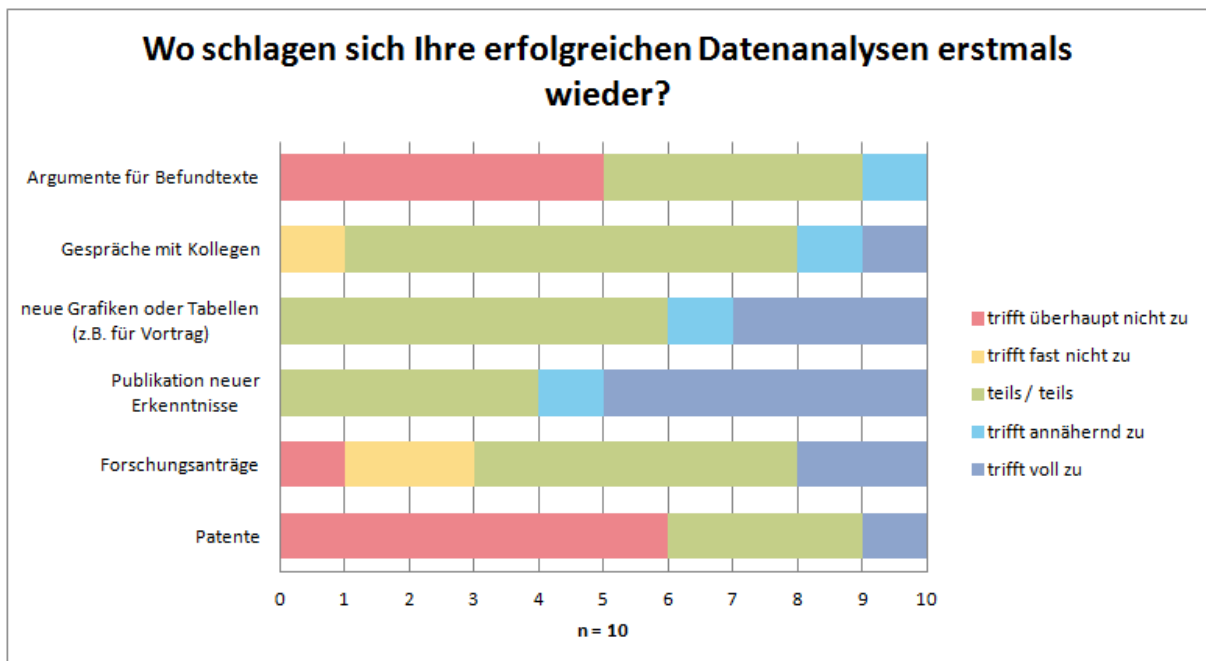
2.5 Arbeitssituation



Datenanalyse findet überall statt, bei der Arbeit, zu Hause und unterwegs – und das zu allen Zeiten. Dennoch gibt es klare Schwerpunkte: räumlich der Arbeitsplatz und zeitlich eher gegen Feierabend und in der Freizeit – also vermutlich außerhalb der Routinen am Arbeitsplatz. Alle praktizierenden Mediziner gaben in der Befragung an, auch in der Freizeit Daten zu analysieren, im Gegensatz zu den Befragten aus den Professionen Medizininformatik/Pharmazie. Unerlässliches Hilfsmittel ist der PC, auch die Papierform findet noch Anwendung. Eher für kooperative Arbeitstechniken geeignete Instrumente wie Beamer oder Whiteboard wurde von niemandem als eingesetzte Hilfsmittel genannt.

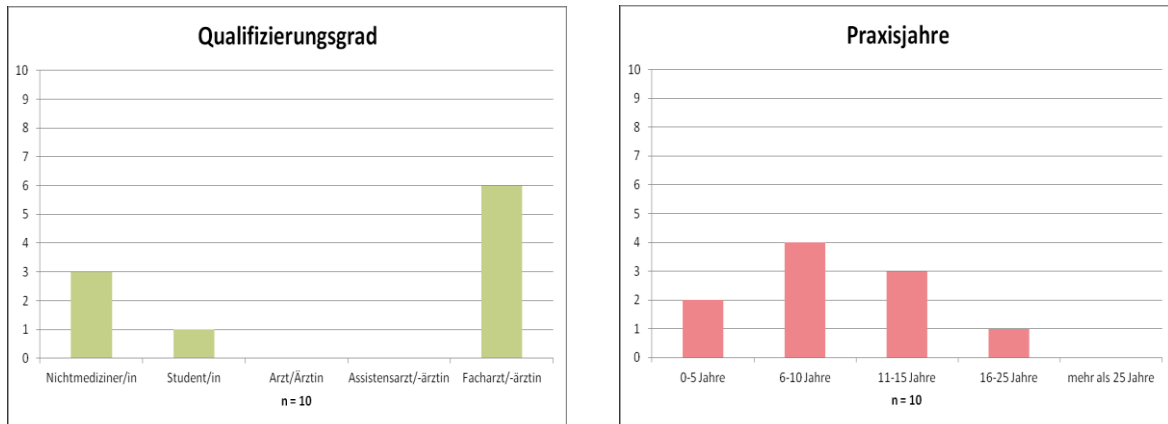
Diese Antworten sprechen für eine räumlich unabhängige, also webbasierte DCT-Entwicklung ohne hohe Zugangshürden. Somit müssen Fragen der Zugriffsrechte geklärt sein und für Berechtigte sollten möglichst geringe technische Zugangshürden bestehen. Die Bildschirmdarstellung der DCT-Entwicklung auch für mobile Anwendungsgeräte wäre wünschenswert.

2.6 Verwendung der Analyseergebnisse



Bei den Antworten auf die Frage, wo sich Ergebnisse von Datenanalysen erstmals niederschlagen, liegt die formale Publikation klar vorne, gefolgt von Ausarbeitungen für Vorträge. Gespräche mit Kollegen über Datenanalysen spielen eine Rolle, die allerdings nicht zentral zu sein scheint. Auf Basis der geführten Interviews lässt sich hier ergänzen, dass persönliche Gespräche mit Fachkollegen (Face to Face oder Telefonat) erste Wahl sind, wenn es um konkrete Fälle geht. Sie finden u.a. im Kontext von Weiterbildungen auf Tagungen oder Arbeitskreisen und der praxis- oder abteilungsinternen Kommunikation statt. Breit gestreut sind die Antworten in Bezug auf Forschungsanträge und Befundtexte sowie Patente. Datenanalyse wird offenbar primär zur Forschung betrieben – also um Antworten zu finden bzw. zu begründen. Problem- oder Hypothesenfindung aus Datenanalysen könnte über Anregungen zur direkten Kommunikation, die das unfertige, suchende einschließt, gestärkt werden.

2.7 Professioneller Hintergrund der Befragten



Zur Bewertung der Antworten haben wir die Befragungsteilnehmer nach ihrem professionellen Hintergrund gefragt. Nur zwei der acht Teilnehmer sind Einsteiger in ihrem Berufsfeld. Sechs der Befragten gaben an, über die Qualifikation als Facharzt/Ärztin zu verfügen. Als Fach- oder Spezialgebiet wurde neben Nephrologie und Innerer Medizin und von Nichtmedizinerinnen Medizininformatik und Pharmazie genannt. Die Fallzahlen sind zu gering, um Unterschiede im Umgang mit Daten je nach Praxiszeit/Qualifizierungsgrad festzustellen.

Aus den Angaben zur Qualifizierung schließen wir, dass die Befragungsteilnehmer über weitreichenden und größtenteils langjährige Erfahrung im Umgang mit medizinischen Daten verfügen. Bei der Auswertung der Antworten erscheint in einigen Fällen eine Differenzierung nach den Gruppen Mediziner/Nichtmediziner erforderlich, da bspw. Fragen, die auf die direkte Behandlung von Patienten zielen von den Nichtmedizinerinnen bereits aus ihrer Rolle heraus als nicht Relevant eingestuft werden.

3 Zusammenfassung und Diskussion

3.1 Ergebnisse

- Eine (eigene) Idee bietet am ehesten Anlass für eine Datenanalyse, Ideensuche dagegen kaum.
- Datenzentren sind für die Befragten eine wichtige Datenquelle.
- Datenanalyse findet unter Einbeziehung einer breiten Wissensbasis statt: publizierte Wissenschaftliche Literatur und Internetrecherchen zu aktuellen Entwicklungen werden eingesetzt.
- Die Haltung gegenüber Komplexen Visualisierungen ist gespalten zwischen Ablehnung und Bevorzugung.
- Die eingesetzten Hilfsmittel deuten auf eine individuelle Arbeitsweise bei der Datenanalyse.
- Datenanalysen finden zumeist am Arbeitsplatz und eher gegen Feierabend statt.
- die Datenanalyseergebnisse finden primär in Publikationen Verwendung, weniger für Forschungsanträge.

3.2 Interpretation

Datenanalysen finden primär zielgerichtet, d.h. hypothesengeleitet, statt.

Sie sind ein individueller Vorgang, bei dem traditionelle Arbeitsweisen bevorzugt werden. Die Analysen werden weniger explorativ sondern eher ergebnisorientiert in Hinblick auf belastbare Aussagen, z.B. in Publikationen, durchgeführt. Das Potential der ausgeprägten Datenaffinität der Befragten wird z.T. durch technische bzw. administrative Barrieren begrenzt.

3.3 Schlussfolgerungen für DCT-Entwicklung

DCT sollte möglichst viele niederschwellige Wege zur Exploration der Daten (sowohl Struktur als auch Datenwerte) anbieten. Dabei können Ideen entstehen, die eine intensivere Auseinandersetzung mit den verfügbaren Daten induzieren.

Eine enge Anbindung an Datencenter wie OpEN.SC mit aufbereiteten und anonymisierten Daten ist sinnvoll, da hier die Bereitschaft zur Nachnutzung am höchsten ist.

Das System sollte die Möglichkeit bereithalten, bei Bedarf weitere Datenzentren zu integrieren. Die DCT sollten eine möglichst breite Auswahl an sekundären Informationsquellen integrieren und deren Nutzung anregen, um die Wissensbasis zu optimieren. Komplexe Visualisierungen sollten eingebunden werden, um der Komplexität der Daten gerecht werden zu können. Gegenüber diesen Visualisierungsanwendungen bestehen zum Teile starke Vorbehalte. Sie sollten daher wahlweise genutzt werden können und keineswegs das zentrale oder einzige Instrument zur Datenexploration in den DCT darstellen. Das Gespräch mit Kollegen nimmt eine untergeordnete Stellung ein. Ein Grund dafür könnte sein, dass sich Rohdaten und deren Kontext verbal meist nur schwer vermitteln lassen. Das Gespräch, so unsere Vermutung, findet erst dann statt, wenn die Daten aufbereitet wurden, z.B. in Form eines Diagrammes. Für die DCT-Entwicklung bedeutet diese Annahme, dass von umfangreicheren Maßnahmen zur Unterstützung der Kommunikation, wie beispielsweise der Implementierung von Social Media, abgesehen werden kann.

Dennoch sollten Möglichkeiten berücksichtigt werden, die die Nutzung von Wissensressourcen durch persönlichen Austausch auch auf dem analogen Weg stärken. So könnten bspw. einfache wie komplexe Visualisierungen die per Beamer projiziert werden, kooperatives Arbeiten erleichtern.

3.4 Informationswissenschaftliche Diskussion der Ergebnisse

Neben dem unmittelbaren Nutzen für die DCT-Entwicklung sind die Befragungsergebnisse für uns auch von allgemeinem informationswissenschaftlichem Interesse. Einige dieser Punkte diskutieren wir im Folgenden.

Zu den Fragen, warum PubMed eine herausragende Rolle bei den Informationsmitteln spielt und andere Instrumente, wie bspw. Google Scholar als Nachweisinstrument für wissenschaftliche Publikationen für unsere Befragten nur sekundären Status besitzt und das persönliche Gespräch mit den Fachkollegen nur im Mittelfeld unserer Umfrage rangierte, könnte die Betrachtung von *Zipfs Principle of Least Effort*⁴ (PLE) von Nutzen sein. Diese Theorie geht davon aus, dass Individuen bei allem was sie tun, möglichst wenig Aufwand betreiben. Die Zielerreichung einer Aufgabe unterliegt demnach der Abwägung einer individuellen Kosten-Nutzen-Rechnung.

⁴ Siehe dazu: Donald O. Case (2005): Principle of Least Effort. In: Fisher, Karen E.; Erdelez, Sandra; McKechnie, Lynne (Hrsg.): Theories of Information Behavior. Cambridge, Mass. : MIT Press. S. 289-292.

Die Umfrageergebnisse sowie die Interviews zeigten, dass PubMed als primäres Recherchemittel dient. Diese eindeutige Präferenz kann zu einem Problem für den Nutzer werden, das ebenfalls im PLE begründet liegt. Glaubt er nämlich mittels PubMed auf ein ausreichendes Rechercheergebnis gestoßen zu sein, fehlt der Anreiz, eine strukturiertere Literaturrecherche, die andere Ressourcen mit einschließt, durchzuführen und damit die Qualität der Rechercheergebnisse zu verbessern. Wir haben nur nach den Vorlieben bei den Rechercheinstrumenten gefragt, nicht nach Art und Weise ihrer Nutzung. Die eindeutige Bevorzugung eines Instruments (PubMed) könnte weiterhin ein Hinweis darauf sein, dass eine unzureichende Wissensbasis genutzt wird, obwohl die Forscher subjektiv den Eindruck haben, alles zu finden, was sie zur Interpretation der Daten benötigen.⁵ Sollte diese Annahme für unseren Fall zutreffen, wäre dies zum einen ein Problem für die Forschung und zum anderen eine Herausforderung an das Informationssystem, dass eine breitere Abdeckung möglicher Quellen bei der Recherche zumindest anregt.

Auch auf die Frage, warum das Gespräch mit den Kollegen als externe Datenquelle nur ab und zu in Anspruch genommen wird kann Zipfs PLE einen Erklärungsansatz liefern. Fließen nämlich die am Häufigsten genannten Antworten zu den zeitlichen (eher gegen Feierabend) und räumlichen (am Arbeitsplatz) Bedingungen für die Datenanalysen mit in die Betrachtung hinein, lässt dies den Schluss zu, dass auch hier individuelle Kosten-Nutzen Abwägungen getroffen werden, z.B. die für den medizinischen Arbeitsalltag durchaus vorstellbare Überlegung, ob dem Kollegen zuvor umfangreich der Kontext für das Recherchevorhaben erläutert werden muss – dazu muss der Kollege auch zeitnah greifbar sein, wenn der Impuls ausgelöst wird – oder die für diesen Vorgang aufgewendete Zeit unmittelbar in die eigene Recherche investiert wird. Auf Zeitpunkt und Umgebung der Datenanalysen kann ein herkömmliches webbasiertes Informationssystem jedoch kaum Einfluss nehmen.

Raya Fidel empfiehlt zur Frage der besseren Integration von Informationssystemen in den Arbeitsalltag den Ansatz der *Cognitive Work Analysis*.⁶ Die Verfolgung dieses Ansatzes könnte auch für die DCT-Entwicklung Perspektiven aufzeigen, wie der Wissensaustausch unter Kollegen besser in datenorientierte Forschungsprozesse

⁵ Vgl.: Lawrence, David W. 2008. The information-seeking behaviors of professionals and information sources in the field of injury prevention and safety promotion. Stockholm: Division of International Health, Department of Public Health Sciences, Karolinska institutet. S. 83f.

⁶ Fidel, Raya (2012): Human Information Interaction: An Ecological Approach to Information Behavior. Cambridge [u.a.] : MIT Press. S. 225f.

eingebunden werden kann. Der Re-Use von Daten spielt für unsere Befragten eine nachgeordnete Rolle. Diese Beobachtung ist möglicherweise für die Medizinische Forschung insgesamt charakteristisch. Wichtige Basis für Re-Use sind leichte Zugänglichkeit, Zuverlässigkeit der Nutzer und Systeme sowie Anerkennung der Leistung für denjenigen, der die Daten für den Re-Use bereitstellt.⁷

3.5 Fazit

Für die Entwicklung der DCT liefert die Auswertung der Erhebungen zahlreiche Anregungen, die in die Gestaltung der Anwendungskomponenten einfließen.

Für explorativen Zugang zu den Daten werden zwei Wege angeboten: Visualisierung mit SemaVis⁸ und Browsen durch die Datenfelder (DCT-Komponente:

Propertybrowser). Damit soll ungerichtetes Erkunden der Daten angeregt werden.

Die DCT-Komponente *Research Field Explorer* führt gleichzeitige Suchanfragen auf verschiedenen Informationssystemen aus, zunächst dem OpEN.SC-Datencenter,

PubMed und Google. Durch die Architektur der Komponente können perspektivisch weitere Informationssysteme zur Ausweitung der Wissensbasis einfach eingebunden werden.

Der Research Field Explorer unterscheidet sich von herkömmlichen Metasuchen durch ein semiautomatisches Mapping auf Standardvokabular⁹, um weitere Daten- und Informationsquellen einzubinden und sprachübergreifend zu recherchieren. Der Zugriff auf die Forschungsdaten soll möglichst ohne hohe technische oder administrative Hürden erfolgen können. Die DCT werden deshalb als Bestandteil des OpEN.SC entwickelt, um auf das dort implementierte und bei den Nutzern eingeführte Rechtemanagement für externen Zugriff auf Daten zurückgreifen zu können.

Dank

Wir danken unseren Interviewpartnern und den Teilnehmern der Onlinebefragung dafür, ihre Zeit zur Verfügung gestellt und uns wertvolle Einblicke in ihre Praxis im Umgang mit Forschungsdaten gewährt zu haben.

⁷ Vgl. Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev U. Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, Mike Frame, und Cameron Neylon. 2011. Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6, e21101. S. 9ff.

⁸ <http://www.semavis.net/>

⁹ Unified Medical Language System (UMLS), <http://www.nlm.nih.gov/research/umls/>

4 Anhang

4.1 Interviewleitfaden

Arbeitsatmosphäre / Arbeitsalltag / Workflow

01 Einstiegsfrage: Woran, wo und wie arbeiten Sie?

Informations- und Kommunikationsverhalten

02 Wann arbeiten Sie mit dem Rechner und welche Kommunikationstechnologien nutzen Sie beruflich, in welchen Kontexten?

03 Wie kommen Sie an die Informationen, die Sie benötigen. Welche Informationsquellen nutzen Sie wann und wozu? (Internet, DB, Buch/Standardwerk, persönliche Gespräche etc.)

Kooperation und Kollaboration

04 Wann sind für ihre Arbeit Kooperationen mit anderen Forschern, klinischer Praxis, externen Partnern wichtig? (Kontexte)

05 Wie offen gehen Sie mit Ideen und eigenen Überlegungen (erste Überlegungen, Skizzen, Gedanken) um und warum?

06 Welche Bedeutung haben für ihre Arbeit Netzwerke, Communities?

Aspekte:

- Größe: Wie groß ist Ihr Netzwerk (Personenanzahl)?
- Kommunikation: Wie kommunizieren Sie und wann?
- Heterogenität: Wie heterogen ist ihr Netzwerk (Personen)?
- Reichweite: Welche Bedeutung hat die nationale und internationale Vernetzung mit Kollegen?

Forschungsdaten

07 Mit welcher Art von Daten/Forschungsdaten haben Sie beruflich zu tun? Wie arbeiten Sie mit Daten (Bilder, Datensätze etc.) und wie kommen Sie an Ihre Daten für Ihre Forschung?

Delegieren Sie die Datenrecherche / Datenauswertung?

08 Wir leben in einer zunehmend datenlastigen Welt. Wie hat sich Ihr Umgang mit Daten innerhalb ihres Arbeitsalltags in den letzten Jahren verändert?

Wie würden Sie sich Ihre Arbeitsumgebung zukünftig wünschen? (Desktop-GUI, Computertools)

Kreativität/Problemlösungskompetenz

I Denken Sie nun mal an Ihr letztes Paper/Studie/Fall: wie sind Sie da, bei der wissenschaftlichen Datenanalyse vorgegangen? Wie sind Sie an Ihre Arbeitshypothese, an ihre Analyseansätze gekommen? Erzählen Sie mal?

Aspekte:

- Welcher *Art* sind die zu lösenden Probleme,
- Welche *Strukturen/Rahmenbedingungen* unterstützen/behindern Problemlösungsprozesse?,

- Zeitdruck vs. ruhige *Arbeitsatmosphäre* (Wann nehmen Sie den *Zeitdruck* aus der Arbeit raus, um ein Problem zu lösen? Nennen Sie Beispiele, in denen Sie Abstand vom Tagesgeschäft genommen haben, um einen Fall/Problem/Fragestellung zu lösen.),
- digitale *Medien*,
- Bedeutung von *Kreativität* für den Prozess der Problemlösung,
- Wann wird kooperiert und wie – auf welchem Weg – , mit wem?, Welche Bedeutung hat der Austausch mit Kollegen (Netzwerk) hinsichtlich neuer Lösungen (Kreativität, Problemlösungen)? Wann ist ein Blick über die Fachdisziplin hinaus wichtig? Wie geht das und wie ist das erfolgreich?
- Wann ist eine Fragestellung ein Problem, das Sie *selber lösen* möchten/ müssen?

II Eine abschließende Frage: Haben Sie Methoden oder Techniken, zur Lösung von Problemen; wie kommen Sie auf neue Ideen und Gedanken?

Soziodemografische Daten

Fachdisziplin/Fachdomäne:

Tätigkeitsschwerpunkt (praktizierend/forschend):

Alter:

Geschlecht:

Ausbildung:

4.2 Online-Fragebogen¹⁰



(1 von 7) Was veranlasst Sie üblicherweise zur Datenanalyse?
(Mehrfachantworten möglich)

Frage zu bestimmten Patient/Fall	Trifft überhaupt nicht zu ☐	Teils / Teils	Trifft voll zu ☐
Forschungsfrage	Trifft überhaupt nicht zu ☐	Teils / Teils	Trifft voll zu ☐
Spontanes, ungerichtetes Interesse an Daten	Trifft überhaupt nicht zu ☐	Teils / Teils	Trifft voll zu ☐
Verfolgen einer eigenen Idee	Trifft überhaupt nicht zu ☐	Teils / Teils	Trifft voll zu ☐
Verfolgen einer Idee aus der Fachliteratur	Trifft überhaupt nicht zu ☐	Teils / Teils	Trifft voll zu ☐

(2 von 7) Welche Datenquellen nutzen Sie üblicherweise für die Analyse?
(Mehrfachantworten möglich)

Patientendaten/Patientenakte	selten ☐	Teils / Teils	häufig ☐
Re-Use der Daten von Kollegen	selten ☐	Teils / Teils	häufig ☐
Zugriff auf Datenzentren	selten ☐	Teils / Teils	häufig ☐
Gezielte Erhebung neuer Daten	selten ☐	Teils / Teils	häufig ☐
	selten ☐	Teils / Teils	häufig ☐

(3 von 7) Welche externen Hilfsquellen ziehen Sie bei der Datenanalyse für gewöhnlich hinzu?
(Mehrfachantworten möglich)

Gespräch mit Kollegen	gering ☐	teils / teils	häufig ☐
Internet (Google u. ä.)	gering ☐	teils / teils	häufig ☐
Google Scholar	gering ☐	teils / teils	häufig ☐
Faktendatenbanken/Nachschlagewerke (Rote Liste, Pschyrembel...)	gering ☐	teils / teils	häufig ☐

¹⁰ Die Teilnehmer konnten bei den Fragen 1-6 auf einer fünfstufigen Skala einen Regler mithilfe von Mausgesten zur für sie passenden Antwort verschieben. Der Regler selbst wurde erst angezeigt, sobald man sich im dafür vorgesehenen Feld mit der Maus befand. Damit sollte eine Beeinflussung des Teilnehmers verhindert werden.

Literaturdatenbank (PubMed-Recherche...)	gering ⌵	teils / teils	<input type="checkbox"/> häufig <input type="checkbox"/> moderat <input type="checkbox"/> selten <input type="checkbox"/> nie
Bestimmte Fachzeitschriften	gering ⌵	teils / teils	häufig ⌵

(4 von 7) Welche Datendarstellungen bevorzugen Sie? (Mehrfachantworten möglich)

Diagramme und Kurve	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵
Komplexe, z.B. interaktive, Visualisierung	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵
Tabelle mit numerischen Werten	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵

(5 von 7) Welche Arbeitssituation liegt vor, wenn Sie eine Datenanalyse vornehmen?

- Ort
- zu Hause unterwegs am Arbeitsplatz
- Zeit
- Eher zu Arbeitsbeginn Eher gegen Feierabend Während Bereitschaftsdiensten In der Freizeit
- Geräte / Hilfsmittel
- PC Beamer Whiteboard Papier/Dokumente

(6 von 7) Wo schlagen sich Ihre erfolgreichen Datenanalysen erstmals nieder? In...
(Mehrfachantworten möglich)

Argumentation für Befundtexte	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵
Gesprächen mit Kollegen	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵
neuen Grafiken oder Tabellen (z.B. Für Vortrag)	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵
Publikationen neuer Erkenntnisse	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵
Forschungsanträgen	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵
Patenten	Trifft überhaupt nicht zu ⌵	Teils / Teils	Trifft genau zu ⌵

(7 von 7) Weitere Angaben zur Statistik

- Qualifizierungsgrad
- Nicht Mediziner/in
 - Student/in
 - Arzt/Ärztin
 - Assistenzarzt/-ärztin

- Facharzt/-ärztin



Fach- oder Spezialgebiet

Dauer der Praxis in Jahren

- 0-5 Jahre
- 6 bis 10
- 11 bis 15
- 16-25
- mehr als 25

Sie haben die Möglichkeit, Ihre E-Mail-Adresse zu hinterlassen, wenn Sie über die Ergebnisse der Befragung informiert werden möchten.