International Journal of Health Services Research and Policy

www.dergipark.org.tr/ijhsrp

IJHSRP

Research Article

CHALLENGES AND FACILITATORS TO THE SECONDARY USE OF ROUTINELY COLLECTED ORAL HEALTH DATA FROM MULTIPLE EUROPEAN COUNTRIES

Heather LEGGETT [1]*   *Frank FOX [2]*   *Karen VINALL-COLLIER[3]*   *Julia CSIKAR[4]*
Evangelos ZORMPAS [5]   *Helen WHELTON [6]*   *Gail VA DOUGLAS[7]*

[1]Honorary Research Fellow. The School of Dentistry, The University of Leeds, Leeds, UK.
[2]Research Fellow. The School of Dentistry, The University of Leeds, Leeds, UK.
[3]Lecturer in Dental Public Health, The School of Dentistry, The University of Leeds, Leeds, UK.
[4]Lecturer in Dental Public Health, The School of Dentistry, The University of Leeds, Leeds, UK.
[5]Research assistant. The School of Dentistry, The University of Leeds, Leeds, UK.
[6]Head of the College of Medicine and Health, Oral Health Services Research Centre and College of Medicine and Health, University College Cork, Cork, Ireland.
[7]Chair in Dental Public Health/Dental Public Health & Health Services Research Theme Lead, The School of Dentistry, The University of Leeds, Leeds, UK.
*Corresponding author: H.leggett@leeds.ac.uk

Abstract: *This research aimed to identify and explore perceived challenges and facilitators to acquiring routinely collected oral healthcare data for research in six European countries with the aim of generating practical solutions for future initiatives. Seventeen participants from the UK, Denmark, Germany, Hungary, Ireland and the Netherlands participated who were either data requestors or data providers for the ADVOCATE project. Focus groups using the nominal group technique were undertaken using PESTLE as a theoretical framework to guide the discussion. The data were analysed using content analysis. Four main challenges were identified: 1) legality rules influencing the data available, 2) variations in data standardization/coding between countries, 3) relationships and responsibilities between stakeholders, and 4) data not available for secondary use. The facilitators included: 1) having a framework in place to guide the process, 2) having strong relationships between stakeholders, 3) having technical elements in place to support the process, and 4) taking a pragmatic approach to the available data. It is hoped that identifying these challenges will raise awareness of potential issues for undertaking such research and that tackling these and building on the facilitators will establish stronger foundations for the sharing of data within and across disciplines and countries.*

Keywords: *Public health, Secondary data analysis, Health information technology, Health services research, Oral health*

1. Introduction

The transition from paper to electronic dental records is accelerating and is driven by the need to modernize our healthcare systems [1]. This development has seen a huge growth in the quantity of clinical data being collected in electronic form [2-4] and presents a significant opportunity for the (secondary use of these data, often termed 'big data'. Big data can be used to inform commissioning decisions, for clinical audit, understand and stratify risk, monitor outcomes and costs, and performance manage payment controls [5]. The EU recognises the value of routinely collected data in medical and clinical research [6] and detailed this in its eHealth Action Plan [7]. Additionally, the EU's study on big

data in public health [6, 8] recommended the expansion of existing sources of big data in health and the promotion of sharing data and improving analytics methods and interoperability. The EU funds many big data initiatives [9] and is also a partner in the Big Data for Better Outcomes [10] initiative, a project which is creating a standardised, federated research data network through the European Health Data for Evidence Network [11]. In the US, the National Institute of Health, launched the 'Big Data To Knowledge' (BD2K) program [12] in 2014. BD2K aims to facilitate the use of biomedical big data, develop and disseminate analysis methods and software, and to establish centres of excellence for biomedical big data.

Against this backdrop, the four-year ADVOCATE (Added Value for Oral Care) project was developed and funded by the European Commission's Horizon 2020 program. Its global aim was to determine how to influence oral healthcare systems towards effective disease prevention, with the aim of making preventive treatments more preferred. Six European countries were partners in the project: The UK (England, Scotland), Ireland, the Netherlands, Germany, Hungary and Denmark. ADVOCATE work-streams implemented a number of parallel approaches to facilitate this change [13]. One such work-stream investigated whether routinely collected oral health data from health insurance companies in the Six European countries could be used to encourage preventive care, as well as highlight best care practices observed in the data. ADVOCATE explored the feasibility and utility of international comparisons between different healthcare systems, both at country level and at the level of individual dentists [13]. Crucially, harnessing the knowledge and value from these datasets depended on successful data acquisition. Negotiating access to data in these countries required a broad understanding of the oral health systems in each location and also necessitated detailed approaches to the various governance systems in place. ADVOCATE encountered many such challenges to its efforts and this case-study aims to enumerate and categorise these using PESTLE [14] analysis as a theoretical framework in the context of prior wider health research. PESTLE focuses on the impact on systems or organisations of six factors: Political, Economic, Sociological, Technological, Legal and Environmental. PESTLE is more often used to analyse external influences on a business [15], but its use within this aspect of the ADVOCATE project was seen to provide a useful framework to classify issues and identify outliers to this framework. Previous research by Van Panhuis and colleagues has explored existing literature to classify the challenges to data sharing and identified six key categories: Political, Economic, Motivational/social, Technical, Legal and Ethical- similar to those used in PESTLE [16]. Later work expanded on the underlying causes of these factors, identifying political, economic and legal obstacles as the most challenging issues to overcome [17]. Consequently, Edelstein and Sane [17] proposed strategies and solutions for these, e.g. trust-building measures, local data governance agreements and data standardisation. Related work in ADVOCATE has proposed process models for acquiring administrative routine data for health services research [18] and alludes to some of the barriers experienced in the data acquisition process. The current research complements this by eliciting and analysing the experiences of a broader range of ADVOCATE stakeholders and relating them to prior similar investigations.

This case-study aims to identify and categorise the challenges encountered by ADVOCATE participants when acquiring routinely collected oral healthcare data for research in six European countries and explores utilised facilitators or potential solutions to these in the data acquisition process. It is anticipated that this will generate fresh insights and practicable learnings applicable to future initiatives.

2. Method

2.1. Participants

Participants were recruited using purposeful sampling, selected due to their role in data acquisition or data provision for the ADVOCATE project across the six partner countries.

Ethical statement

Ethical approval was granted from the Dental Research Ethics Committee at the University of Leeds (Date: 5 April 2017; Number: 051115/HL/182).

2.2. Design

The key step in this study was a focus-group workshop using the Nominal Group Technique (NGT) to explore participants' perceptions of the main challenges and facilitators (C&F) to data acquisition and its usage (Figure 1). Prior to the workshop, a pre-study phase was undertaken through participant online engagement in a tool called Well Sorted® (Step 1); participants were invited to individually give their ideas, thoughts and comments on data acquisition and its usage. This enabled us to gain preliminary insights of the participants' experiences of acquiring or providing data. This pre-study phase helped to develop a qualitative topic guide to support the focus group discussions at the NGT focus-group workshop alongside the PESTLE framework (Step 2). This led to the identification of participant's 3 top C&F's (Step 3). The workshop was followed by re-engagement with Well Sorted® two months later to provide verification of the C&Fs identified by participants at the workshop (Step 4).

The techniques used in this study are described below:

PESTLE: The modified version of the PESTLE technique was used as a theoretical framework to guide the exploration of the challenges and facilitators to data acquisition and its use[16]. In line with Van Panhuis' approach we omitted the 'Environmental' factor and instead used the 'Ethical' factor. We believed that the sharing of computer based oral healthcare data between providers and countries would have a greater ethical rather than environmental impact. As such, asking participants to discuss the ethical aspects of data sharing was more applicable/meaningful to our research aims than discussing the environmental aspects.

Nominal Group Technique (NGT): A structured approach to group decision-making using focus groups. It fosters the generation of ideas and supports equal group-member participation to produce a rank-ordered set of decisions or outcomes [19, 20]. Participants are asked questions in small groups by a moderator to generate ideas as individuals. The ideas are then prioritised in order of importance within the group. NGT is recommended when group consensus is important, in particular when there are several similar ideas or suggestions [21].

Well Sorted®: An online resource which was used to facilitate participant led, thematic development of ideas and to verify the findings from the focus group workshop through triangulation.
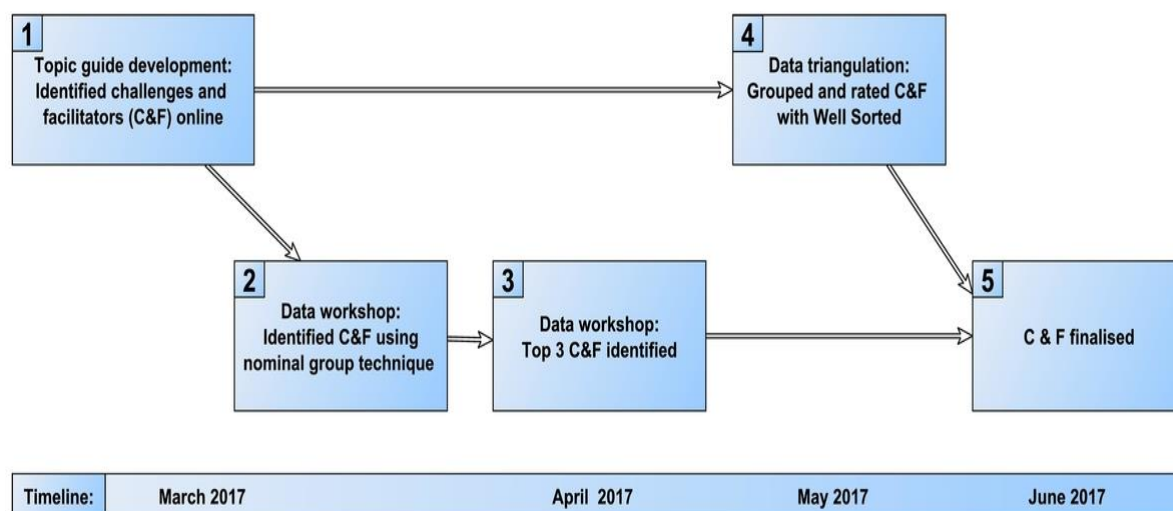
Figure 1. Diagram of the procedure

2.3. Procedure

Ethical approval was granted from the Dental Research Ethics Committee at the University of Leeds (051115/HL/182). Participants were sent an information sheet to read before participating and asked to sign the consent form if they agreed to participate.

Of 20 invited participants, 17 agreed to participate and attended the workshop (Table 1). All participants were involved in the data acquisition process for ADVOCATE. Participants represented four groups: (1) organisations who held/owned the data (data provider), (2) those involved in requesting data from a data provider (data requestor), (3) those involved in organising, consolidating and analysing data collected (data analyst), or (4) those hosting the data on a secure platform (data host). The data providers were from private insurance companies (Germany, Ireland and the Netherlands) or national insurers (England, Scotland, Denmark, Hungary). The data requestors for Denmark, Ireland, Hungary, and Germany were researchers from the ADVOCATE project. ADVOCATE researchers from the University of Heidelberg (Germany) acted as data requestors for data from Scotland and the Netherlands as well as the data analysts for the project.

A focus-group workshop was held in April 2017 with the 17 participants (Table 1) where they were spilt into three groups. Each group had two researchers (not involved in the data acquisition process); one led the discussion and the other took notes and moderated the meeting; this ensured that the discussion stayed on track and that everyone had the chance to participate. Each participant was asked to talk through their experiences of where the C&F occurred during data acquisition and/or provision. A top-down, theory driven approach was adopted using the modified PESTLE technique with the topic guide, to facilitate the discussion and ensure that all areas received consideration. This created group discussion between participants on the key C&F (Figure 1. Step 2). The focus groups were audio-recorded and the C&Fs discussed were captured on post-it notes during the session to enable the participants to see what had been discussed. The participants were also encouraged to note any thoughts or ideas on their own post-it-notes. After discussing their thoughts around C&Fs to data acquisition/provision, the participants were asked (within each of the three groups) to agree on what they saw to be the three main challenges and three main facilitators/solutions to data acquisition.

The nine challenges and nine facilitators/solutions identified through the three group discussions were presented to the whole group. Participants picked their top three challenges and top three facilitators and those with the most votes were named the most important/influential. The sub-groups

came back together and reviewed the top 9 C&Fs and through a group discussion consensus was reached on the top 3 C&Fs (Figure 1. Step 3).

Table 1. Demographics of participants

Country	Number of participants	Organisation	Role
The UK: England	3	NHS England n=1	Requestor
		NHS Business Service Authority (National insurer) n= 2	Provider
The UK: Scotland	3	NHS Scotland n=1 (National insurer)	Provider
		Aridhia n=2	Data host
Netherlands	1	Achmea (Private insurer)	Provider
Hungary	1	Semmelweis University	Requestor
Germany	5	University of Heidelberg n=4	Requestor/ Data analyst
		SpectrumK n=1 (Private insurer)	Provider
Denmark	1	University of Copenhagen	Requestor
Ireland	3	University College Cork n=2	Requestor
		Decare (Private insurer) n=1	Provider

2.4. Analysis

The audio recordings, notes taken and post-it notes from the focus groups were analysed using content analysis [22] to describe each C&F based on the participants' experiences. C&Fs were explored across countries rather than on a country level. This was to ensure anonymity and because most challenges and facilitators were present across countries rather than being specific to any one country. Categorization matrices are used to organise qualitative data when undertaking content analysis. The identified C&Fs were utilised as categories in a primary categorization matrix around which the audio recordings, notes and post-it-notes were analysed [23]. The PESTLE categories were also used as a secondary categorization matrix and were mapped onto the C&F. In this instance, each identified C&F was a category in the table illustrated by a row. The data from the audio recordings, notes and post-it notes were reviewed for content and coded into columns for the corresponding category with exemplification of that category through the identification of phrases, sentences or words which fitted under the categorization matrix and were coded accordingly. Any data which did not fit the categorisation matrix was held separately and analysed for relevant additional information. After the categorisation matrix was populated the analysis became iterative and the categories underwent constant refinement as more data were read and analysed. The notes and post-it notes were analysed by one researcher and checked with another two. Any differences were resolved through discussion within the research team. As the detailed notes and the information in the post-it notes were comprehensive, the audio recordings were not transcribed but referred to if clarification or extra information was required on any points raised by the participants.

Data triangulation is often used in qualitative research as a method of validating findings by comparing data from different sources or approaches at different times and places, or from different people [24]. We used Well Sorted® as a method of triangulating our findings from the group workshop with the same participants two months later. At this point, participants were re-sent the list of important C&F they had generated in the topic guide development task using Well Sorted®. The participants were asked to sort all these items into groups of similarity. They were asked to complete this twice, once focussing upon the challenges items and once again for the facilitator/solution items. One set of groups was created for the challenges items and another set was created for the facilitator/solution items. The groups that were created via Well Sorted® approximated the C&F identified from the workshop and were included in the analysis.

3. Results

Four main challenges and four main facilitators/solutions were identified (Table 2). The challenges were mostly political, with one technical and one social/motivational barrier. Ethical issues were touched upon but within the content of other issues. Facilitators were largely technical and social/motivational. The C&Fs are discussed below in more detail; their corresponding PESTLE code is found after the title in brackets. Exemplar quotes placed under each category from the analysis can be found in Appendix 1.

Table 2. Challenges and facilitators to data acquisition

Challenges to data acquisition
Legality, including privacy rules influencing data available (Political/legal)
Variations in data standardisation and coding between countries (Technical)
Relationship and responsibilities between stakeholders (Social/motivational)
Data is collected but not made available for secondary use (Political)
Facilitators to data acquisition
Having clear detailed roles within the data collection process, rules and a framework in place   to guide the process (Technical / Social/motivational)
Having strong relationships between stakeholders (Social/motivational)
Technical elements in place to support the process (Technical)
Taking a pragmatic approach to the available data (Social/motivational)

3.1. Challenges

3.1.1   Legality, including privacy rules influencing data available (Political/legal)

Legal restrictions focusing around privacy rules were a substantial barrier to obtaining and sharing routinely collected dental activity data. The restrictions prevented data providers from sharing the requested data or from allowing direct access to the data, it also resulted in anonymising the data to a point where it lost its value. The requirement of a third party for anonymisation caused issues, as did ongoing legislative changes. In one case, pre-2006 data became unavailable when the commissioning system changed how dental activity was recorded and reimbursed. Being able to compare treatment patterns before and after this policy change would have been valuable. In another case, it was not always possible to link up patients' information, this hindered more detailed analysis due to the low level of data provided.

3.1.2   Variations in data standardisation and coding between countries (Technical)

Comparison of datasets was often difficult due to lack of a consistent, harmonised data collection and storage system between countries. These factors made it difficult for those analysing the data to conduct effective cross-country comparisons. A lack of clarity among some data providers regarding which variables were required and which were available hindered the data sharing process. Navigating one data authority's website (responsible for permissions) was reported to be difficult. Technical matters were not always defined in advance leading to problems with data supply/transfer, software compatibility or hardware requirements.

3.1.3   Relationship and responsibilities between stakeholders (Social/motivational)

There were difficulties building a rapport between the data providers and requestors when communication was undertaken via email. Staff turnover in some organisations was high and this negatively impacted relationship continuity; frequently 'old ground' had to be covered again with each new individual. Relationship development was often needed which proved time-consuming. Some

providers queried the benefits to them of giving their data to the ADVOCATE project. One felt that demands were placed on overworked staff with low motivation, contributing to delays.

3.1.4   Data are collected but not available for secondary use (Political)

Data are most often collected for administration purposes not for research. Participants felt that this exacerbated the issues surrounding how and why data were collected, coded and stored. Thirteen of the participants explicitly shared views that they believed it was vital to lobby for legislation mandating that routine data be made available for health improvement research. Furthermore, private health insurance companies are not obliged to deliver data on health or health services and therefore need a clear incentive to do this.

3.2. Facilitators

3.2.1   Having clear detailed roles within the data collection process, rules and a framework in place to guide the process (Technical /Social/motivational)

A number of suggestions revolved around organisational matters such as roles, standards and procedures. Suggestions were given for how a framework could be developed to make the technical side of the process simpler in the future including the use of standard coding systems such as SNOMED/SNODENT [24]. Agreeing the required data elements and their formats in advance was seen as a solution to some of the issues surrounding getting different levels of data in different countries. A harmonisation table was created by data requestors to allow comparison of dental care indicators, based on claims codes from different countries. The use of standard procedure codes could allow for more accurate comparison of their own data with peer data. One provider felt confident that they could provide data for several of the indicators in the harmonisation table. A practical manual developed by the data handlers and requestors from the ADVOCATE team was highlighted as important as it helped to offer guidance and a basis for communication.

3.2.2   Having strong relationships between stakeholders (Social/motivational)

A strong relationship with the others in the process helped to facilitate the acquisition of the data. Having face to face meetings with colleagues was seen as one way to build this working relationship. A secure online research workspace and online data platform was utilised which was seen as strengthening the trust between the requestors and providers. Another facilitator was the providers' experience with the specific claims data. It was seen to make things easier as they would be knowledgeable on data quality, the variables and technical processing of the data, and were more likely to have the necessary trained staff. Similarly, the researcher having experience of the data was also a facilitator where they could more easily predict and resolve potential problems in advance.

3.2.3   Technical elements in place to support the process (Technical)

Participants reported on technical elements supporting the process and also suggested some that would aid the process further. Detailed data agreements should be in place regarding elements such as sample size, variables provided, aggregation level and any data protection regulations. Legal and ethical issues should be researched in advance to expedite releasing the data.

It was thought that it would be advantageous to have a central data repository which would cover the provision of public data to third level institutions for research purposes. Participants suggested that raising awareness of the importance of using data within research to increase and encourage the availability of data for research purposes would also be advantageous. Having a data controller with a standard process for giving data access approval which balanced the public health benefit with privacy concerns was also considered to be an important element.

3.2.4   Taking a pragmatic approach to the available data (Social/motivational)

The data analysts and requestors found that a useful approach was to be pragmatic with what they could get, accepting what could be shared with them, even if it was not what had initially been agreed. This enabled them to progress with their analysis despite the setbacks.

4. Discussion

This research explored the real-life experiences of those involved in requesting or providing healthcare data in the ADVOCATE project. The objective of this case-study was to pinpoint the key C&Fs to successful data acquisition and exploitation. These findings provide preliminary insights into identifying and understanding the perceived C&Fs to using routinely collected oral healthcare data for research in six European countries.

Most challenges were found to be due to Political, Social/Motivational or Technical issues, which were also found to be key categories in earlier research [16]. Issues with the relationships between data providers and data requestors were found to be detrimental to the success of the process. The corollary of this was also true: when technical elements worked well and there were strong working relationships between providers and requestors, the process ran more smoothly. Accordingly, technical competencies and settled mature relationships between the data requestors and providers were key facilitators to data access. Several of the challenges encountered related to the fundamental infrastructure necessary for the sharing of healthcare data. A legislative basis for routine accessing of such data for research was lacking and, in some participants' experience, the existing legal regimes on data protection and privacy were used as justifications for denying or restricting access to the requested data. This agreed with earlier work [25, 26] where it was suggested that if challenges of privacy and data protection can be adequately addressed, oral health related data contained in electronic health records and claims databases could reveal a huge potential in significant clinical knowledge collection and better understanding of patient disease patterns. Furthermore, clear definitions of the employee project roles and their responsibilities were lacking in many cases. This lack of fundamental infrastructure is not surprising given that the secondary use of oral health data for research on the scale undertaken in ADVOCATE has not been attempted before in a European setting. It was suggested that health policies focusing on health rather than finance would help alleviate this issue. Furthermore, political lobbying would be required to highlight the benefits gained from providing the necessary infrastructure and organisational competencies to facilitate this and to put the necessary legislation in place to mandate data sharing. A lack of consistency in the data collected, and their availability in different countries added further complications to the process. In ADVOCATE, the additional complexity of pan-European cooperation, incorporating language, cultural, and legislative variety added to these challenges and further compounded the local challenges of the lack of data standardisation, diverse coding systems and data incompatibility. This key finding - the need for an established research infrastructure - agreed with developments at Vanderbilt University Medical Centre (USA) [1], emphasising that such an infrastructure enabled better access for the research community. The participants prioritised strong policies and legislative measures to mandate collection and sharing of data. Clear definitions of the legal and ethical basis for using the data for research would help facilitate this. Technical preparation and unambiguous data requests would help to ensure that the data were of appropriate detail and quality and that its sharing would be standardised and automated.

Using PESTLE to guide the focus group discussion was a useful scaffold and enabled the classification of the final C&F identified. However, it was found that some items fitted into more than one category; further discussion of these may have defined them more and enabled them to be aligned into one specific category. As outlined in the design section we modified the PESTLE framework to

explore ethical rather than environmental issues, as in previous research [16]. The process we adopted of initial individual idea generation through Well Sorted® to develop the topic guide followed by face-to-face group discussion and concluding with individual sorting of the whole group responses, utilises approaches similar to the Delphi technique [27] and Q-sort methodology [28] and enabled back-checking and triangulation of the results. By allowing the participants to consider the C&F individually before having a group discussion enabled the development of a specific and participant relevant topic guide. It also prompted the participants to think about the C&Fs they experienced and it was hoped to elicit more holistic views and address the potential limitations of 'group think' or 'researcher bias'. This may have facilitated a more detailed/in-depth discussion of these factors during the NGT.

The use of individual interviews or less structured focus groups may have led to richer findings with greater in-depth exploration of participant's experiences and attitudes towards data acquisition. However, the method we adopted led to the identification of C&Fs to data acquisition by the group as a whole rather than being drawn out by the researchers during analysis as would be the approach with more traditional qualitative research. A collegiate approach such as this was important as we were exploring data sharing issues across a wide footprint. Although our method is transparent and offers credibility to the findings, future research could look to undertake interviews with participants to explore these C&Fs further. In addition, as we did not transcribe the audio-recordings it is possible that some information was lost between the workshop discussion and analysis. However, we believe that using our detailed notes, the post-it notes created during the workshop and listening to our audio recordings during coding enabled an authentic and robust analysis. Indeed, previous research recognises the limitations of relying solely on audio recordings transcribed verbatim and views the use of written field notes as robust [29]. Furthermore, content analysis was an appropriate analytical technique for our method as it is a technique often used for its wide applicability to the analysis of text from alternative sources as opposed to the more traditional use of interview and focus group transcriptions [23].

The participants were those with recent and intimate knowledge of oral health activity data from six European countries and were in a prime position to discuss and provide insight into the C&Fs of the use of such data for research. However, the generalisability of the C&Fs identified may be a limitation as the participants were discussing their experiences related to data acquisition specifically on the ADVOCATE project. Despite this, the factors identified are similar to those identified in related research and we believe they are likely to be experienced by others requesting or providing data. In addition, whilst we invited the key individuals in each country, there were others involved in the process who were not able to participate. As such, it is possible that vital insights were lost. Interestingly, given the cross-country scope of the ADVOCATE project, language was not seen as a barrier to data acquisition. We believe this is likely due to the fact that many requests were dealt with within each country and any outside-country communication usually involved those in the project who were confident and competent speaking in the English language.

This work is unique and timely as this study brought together private and national dental insurers with data requestors from each of the six countries who had accumulated a wealth of information and experience acquiring data for the ADVOCATE project. To our knowledge, this is the first time data 'providers' and 'requestors' have been brought together internationally to compare challenges to the secondary use of dental activity data. The variety of participants in the workshop proved valuable and resulted in a broad range of issues being discussed, ranging from individuals who saw the necessity for political and legislative changes to those who were more involved with the 'nitty-gritty' of data and its properties. Despite the limitations discussed, this research served to demonstrate the applicability of the approach, techniques, and tools employed in identifying and classifying challenges and facilitators to the acquisition and sharing of international health data, and particularly dental activity data.

5. Conclusion and Future directions

Through active exploration of facilitators and solutions to the challenges encountered, this work expands the previous knowledge in the area which focused on challenges alone. This is in keeping with the EU objectives of making data 'discoverable, accessible, assessable, reusable and interoperable' [8]. and may provide valuable guidelines to similar projects. Given the widespread international interest in exploiting the potential of big data, the findings of this study are important for future work involving the secondary use of routinely collected oral health systems data. Based on our findings we would recommend that stakeholders wishing to use data for secondary purposes continue to lobby for the mandatory collection and sharing of such data and that they simultaneously develop and nurture the relationships between the stakeholders. This will involve clear statements of goals and the creation of 'win-win' scenarios for all parties. The development of trust between the parties is imperative in order to facilitate this. Although this project focused on oral healthcare activity data, the findings are equally applicable across the full spectrum of health data.

The COVID-19 pandemic has generated a strong impetus for facilitating the sharing of data on an international level and future work could also consider innovations as in the EU's EHDEN project [11] and the EVOTION project [30]. EVOTION [30] is seeking to use big-data to inform evidence based health policies within hearing research. Their strong focus on stakeholder engagement will hopefully tackle some of the social and motivational issues highlighted by our participants. EHDEN implements strategies to address technical, legal, and ethical challenges by creating a federated data network using a common data model. Social and motivational issues could be addressed through better communication between the stakeholders, dedicated roles and responsibilities, and the fostering of trust relationships. This could be aided by the formal documentation of organizational structures and face-to-face interaction between stakeholders.

Ethical statement

Ethical approval was granted from the Dental Research Ethics Committee at the University of Leeds (Date: 5 April 2017; Number: 051115/HL/182).

Conflict of interest

The authors have no conflicts of interest.

Authors' Contributions:

H.L: Conceptulised and designed the research, undertook the research and analysis and drafted the manuscript. (100%)

F.F: Conceptulised and designed the research, undertook the research and analysis and drafted the manuscript. (100%)

K.V.C: Conceptulised and designed the research, undertook the research and contributed to drafts of the manuscript. (60%)

J.C: Conceptulised and designed the research, undertook the research and contributed to drafts of the manuscript. (60%)

E.Z: Undertook analysis and contributed to drafts of the manuscript. (40%)

H.W: Conceptulised and designed the research and contributed to drafts of the manuscript. (50%)
G.D: Conceptulised and designed the research and contributed to drafts of the manuscript. (50%)
All authors read and approved the final manuscript.

References

[1]   Danciu, I., *et al.*, "Secondary use of clinical data: the Vanderbilt approach," *Journal of biomedical informatics,* 52, 28-35, 2014.

[2]   Raghupathi, W., Raghupathi, V. "Big data analytics in healthcare: promise and potential" *Health information science and systems,*2(1), 3, 2014.

[3]   Safran, C., *et al.*, "Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper," *Journal of the American Medical Informatics Association,* 14(1), 1-9, 2007.

[4]   Auffray, C., *et al.*, "Making sense of big data in health research: Towards an EU action plan," *Genome Medicine,* 8(1), 71, 2016. /doi: 10.1186/s13073-016-0323-y.

[5]   Border, P., "*Big Data and Public health",* POSTnote, UK Parliment, 2014: Available: http://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-474

[6]   European Commission, "The Use of Big Data in Public Health Policy and Research" 2014.

[7]   European Commission, "eHealth action plan 2012-2020: innovative healthcare for the 21st century," *Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. Brussels,* 2012.

[8]   Habl, C., Renner, A., Bobek, J., Laschkolnig, A. "Study on Big Data in Public Health, Telemedicine and Healthcare" *European Commission*, 2016.

[9]   Pastorino, R., *et al.*, "Benefits and challenges of Big Data in healthcare: an overview of the European initiatives," *European Journal of Public Health,* 29, 23-27, 2019.

[10] Big data for better outcomes., "About BD4BO." https://bd4bo.eu/index.php/about-the-programme/about-bd4bo/ (accessed 4.02.2020).

[11] EHDEN. "European Health Data & Evidence Network." https://www.ehden.eu/ (accessed 04.02.2020.

[12] Bourne, P.E, *et al.*, "The NIH big data to knowledge (BD2K) initiative," *Journal of the American Medical Informatics Association,* 22(6), 1114-1114, 2015.

[13] Leggett, H., *et al.*, "Toward more patient-centered and prevention-oriented oral health care: the ADVOCATE project," *JDR Clinical & Translational Research,* 2(1),5-9, 2017.

[14] Perera, R. The PESTLE analysis. Nerdynaut, 2017.

[15] Basu, R., "*Implementing quality: a practical guide to tools and techniques: enabling the power of operational excellence",* Cengage Learning EMEA, 2004.

[16] Van Panhuis, W.G., *et al.*, "A systematic review of barriers to data sharing in public health," *BMC public health,* 14(1)*,* 1-9, 2014.

[17] Edelstein, M., Sane, J., "Overcoming Barriers to Data Sharing in Public Health: A Global Perspective," *Chathamhouse.org*. [Online]. Available:

https://www.chathamhouse.org/publication/overcoming-barriers-data-sharing-public-health-global-perspective

[18] Haux, C., et al., "A process model for acquiring international administrative routine data for health services research," *GMS Medizinische Informatik, Biometrie Und Epidemiologie (online),* 15(1), 2019.

[19] Harvey, N., Holmes, C., "Nominal group technique: an effective method for obtaining group consensus." *International journal of nursing practice*, 18(2), 188-194, 2012.

[20] Delbecq, A.L., Van de Ven, A.H., "A group process model for problem identification and program planning," *The Journal of applied behavioral science,* 7(4), 466-492, 1971.

[21] Center for Disease Control and Prevention, "Gaining Consensus Among Stakeholders Through the Nominal Group Technique.," 7, U.S. Department of helath and Human Services, 2018. https://www.cdc.gov/HealthyYouth/evaluation/pdf/brief7.pdf

[22] Mayring,P., "Qualitative content analysis," *A companion to qualitative research, 1*, 159-176, 2004.

[23] Elo, S., Kyngäs, H., "The qualitative content analysis process.," *Journal of Advanced Nursing*, 62(1), 107-115, 2008.

[24] Cornet, R., de Keizer, N., "Forty years of SNOMED: a literature review," in *BMC medical informatics and decision making*, 8(1), S2, 2008.

[25] Patti DiGangi, R., "Dental Electronic Health Records: Meaningful and Useful," *RDH,* 2012.

[26] Cantrill, J., et al., "The Delphi and nominal group techniques in health services research," *International Journal of pharmacy practice,* 4(2), 67-74, 1996.

[27] Brown, S.R., "Q methodology and qualitative research," *Qualitative health research,* 6(4), 561-567, 1996.

[28] Mooney, S.J., Pejaver, V., "Big data in public health: terminology, machine learning, and privacy". *Annual review of public health*, 39, 95-112, 2018.

[29] Wengraf, T., *Qualitative research interviewing: Biographic narrative and semi-structured methods.* Sage, 2001.

[30] Saunders, G.H., et al., Application of big data to support evidence-based public health policy decision-making for hearing. *Ear and hearing*, 41(5), 1057-1063, 2020.

Appendix 1. Example quotes for each theme.

Challenges	Quote
Legality, including privacy rules influencing data available (Political/legal)	We experienced limited data access due to data privacy rules P17 DR/DA As a research team we have no direct insight, which data institutions, companies or authorities are collecting, how they are storing them etc. This makes it difficult for us to communicate exactly what data we would like.P2 DR/DA
Variations in data standardisation and coding between countries (Technical)	Plethora of coding systems leads to mapping issues. P1 DR Lack of standardisation in terminology for key data P8 DP If the technical details of data supply/transfer are not clarified in advance, there might be problems with software (compatibility) or hardware (memory requirements). P5 DP Treatments are recorded differently by each country, which makes comparison difficult. P14 DR
Relationship and responsibilities between stakeholders (Social/motivational)	Being unable to discuss the data requirements in person, and instead via email makes it harder to discuss what data [*department name*] can provide and how best it can be used to meet your requirements. P6 DP Permissions and application is often time consuming and practically difficult to complete (not the content) and it may be difficult to get access to a person who can advise in case of questions P7 DR It is not always clear what they want. This makes it harder for us to provide it P16 DP
Data is made available for secondary use (Political)	Private insurance companies are not obliged to deliver data on health or health services P7 DR The Data Protection Commissioner prohibits the use of individual level data which has not been anonymised. Unfortunately the data owner does not have anonymised data so a third party, approved by the DPC, is required. This can lead to previously held assumptions being no longer valid and as a result unforeseen delays with data acquisition, including the need for Data Supply Agreements etc. P9 DR Lack of awareness of the importance of data-research (health service research, epidemiology etc.) P11 DR

Facilitators	Quote
Having clear detailed roles within the data collection process, rules and a framework in place (Technical /social/motivational)	At the time of writing the proposal there must be a confirmation to provide data. This confirmation must be as detailed as possible: -sample size, -provided variables, -time period, -aggregation level, -data protection regulations, -technical details P5 DR We developed a manual to guide the process P4 DR/DA Research all the legal and ethical issues carefully and summarise them for the compliance officer so that he / she is able to answer questions for senior management or board of directors so that data can be released promptly P7 DP Utilise the existing Internationally recognised SNOMED CT which incorporates a dental sub set of coding. This would eliminate need for mapping and concentrate more on data sets. P1 DR
Having strong relationships between stakeholders (Social/motivational)	Researchers having expertise with secondary data. He knows possible data provider and their data (suitable for special research question) and is known by the providers. He knows about emerging problems with gathering/analysing claims data and can clear problems/ask crucial questions in advance. P5 DR Having face to face meeting with all involved. P4 DR/DA
Technical elements in place to support the process (Technical)	We have a standard process for giving approval for this type of data which balances the public benefit against privacy concerns. P6 DP When getting historical data from remote location make sure it is done correctly and accurately on one occasion, no double dipping- it significantly delays progress of study.P7 DP
Taking a pragmatic approach to the available data (Social/motivational)	We adopted a very flexible and pragmatic approach and just accepted whatever data the owners could share with us. P2 DR/DA Make it mandatory that all health service data can be obtainable (of course anonymously) P11 DR

Key:
Data provider = DP
Data requestor= DR
Data analyst =DA