Canadians' Views about Using Big Data in Health Research from a National Online Survey: A Partnership of Patient-Consumers and Researchers

Natalie McCormick[1], Clayon Hamilton[1,2], Cheryl L. Koehn[3], Kelly English[4], Allan Stordy[5] and Linda Li[1,6], [1]Arthritis Research Canada, Richmond, BC, Canada, [2]Physical Therapy, The University of British Columbia, Vancouver, BC, Canada, [3]Arthritis Consumer Experts, Vancouver, BC, Canada, [4]Arthritis Patient Advisory Board, Arthritis Research Canada, Richmond, BC, Canada, [5]Canadian Skin Patient Alliance, Calgary, AB, Canada, [6]Department of Physical Therapy, The University of British Columbia, Vancouver, BC, Canada

Background/Purpose:

Findings from health research using "big data" (large sets of routinely-collected healthcare data) have benefitted individual arthritis patients and society as-a-whole. However, growing public concerns about personal information being accessed for unintended purposes could erode trust in this research and its potential benefits.

We conducted this study to elucidate the public's views on the role of big data in health research.

Methods:

This was a partnership between researchers at Arthritis Research Canada and consumer-patient leaders from three joint and skin disease patient organizations. We developed an online survey assessing: 1) Perceptions about the role of big data in health research and access and privacy controls, 2) Willingness to participate in projects using big data, 3) Major concerns, and 4) Interest in learning more. Respondents were recruited via the websites, e-mail lists, and social media channels of Canadian health research groups, including the three patient organisations.

Survey: Our three-part, ~20-min. survey was open from Jan-Aug 2017. Part 1 asked about respondents' familiarity and initial perceptions about using big data in health research. In Part 2 they were provided with some background information, then asked their views on specific topics (i.e. benefits of using big data, data access and privacy). In Part 3 respondents were queried further about their perceptions of big data and ongoing educational needs.

Analysis:

For each question, we calculated the percentage of respondents selecting each response option.

Results:

151 individuals completed the survey (117=77% female; 47% aged 50-69 years, 28% aged 30-49 years). 101 (67%) had a chronic disease.

At the start of the survey (Part 1), 79% of respondents felt positively about the use of big data for health research and 95% knew the term "electronic health/medical record", but only 58% knew the terms "administrative data/health database".

In Part 2, respondents felt the ability to study large numbers of people (selected by 73%) and long-term effects and rare events (76%) were the top benefits of using big data; long-term treatment effects and disease complications were the most important research topics **(see Table)**.

59% felt the use of big data should be approved by university research ethics boards, and 67% wanted to learn more about how data stewards grant access to data. De-identifying personal information was the most important privacy measure (selected by 89%).

At the end of the survey (after viewing background information about big data), 93% felt positively about big data (vs. 79% at the start), but only 58% were confident about privacy and security measures in place.

Conclusion:

While ethics board approvals and de-identification of healthcare data were highly regarded, more public education, especially about data access and privacy controls, may enhance public trust about using big data in health research.

Table: Percentage of Respondents Selecting Each Feature and Data Access/Privacy Measure Described in Part 2 of the Survey (Most- to Least-Frequently Selected)							
Advantages of Using Big Data for Health Research							
Most Important *(select up to 3)*	**1. Long-Term Effects and Rare Events**	**2. Large Numbers**	**3. Study Potentially-Harmful Treatments**	**4. General Population Comparisons**	**5. More Inclusive**		
	75.5%	72.8%	50.3%	46.4%	43.0%		
Want Additional Information About *(select up to 3)*	**1. Study Potentially-Harmful Treatments**	**2. Long-Term Effects and Rare Events**	**3. Large Numbers**	**4. General Population Comparisons**	**5. More Inclusive**		
	64.2%	63.6%	44.4%	41.7%	35.1%		
Health Research Questions to Study Using Big Data							
Most Important *(select up to 3)*	**1. Treatment Benefits**	**2. Treatment Harms**	**3. Disease Complications**	**4. Changes in Policy or Practice**	**5. Quality of Care**	**6. Cost-Effectiveness**	**7. Risk Factors for Disease** ¦ **8. Disease Incidence and Prevalence**
	55.6%	55.0%	52.3%	43.7%	30.5%	27.8%	23.2% ¦ 9.9%
Data Access Controls							
Most Important *(select up to 2)*	**1. Must Apply for Data Access**	**2. Approval from Research Ethics Board**	**3. Approval from Data Stewards**	**4. Access Data for Limited Time**			
	62.3%	58.9%	51.0%	20.5%			
Want Additional Information About *(select up to 2)*	**1. Approval from Data Stewards**	**2. Approval from Research Ethics Board**	**3. Access Data for Limited Time**	**4. Must Apply for Data Access**			
	66.9%	46.4%	29.8%	25.2%			
Privacy and Security Controls							

Most Important *(select up to 3)*	1. Data are De-Identified	2. Privacy Training and Confidentiality Agreement	3. Review of Research Outputs	4. Funding Agencies Cannot Access Data	5. No Access Outside Canada	
	89.4%	57.6%	43.7%	35.8%	35.1%	

Expressed as the percentage selecting each response option; as multiple responses could be selected, the sum of percentage-frequencies exceeds 100%

To cite this abstract in AMA style: