

Lumos: a statewide linkage programme in Australia integrating general practice data to guide system redesign

Patricia Correll,¹ Anne-Marie Feyer,² Phuong-Thao Phan,¹ Barry Drake,² Walid Jammal,³ Katie Irvine,¹ Adrian Power,¹ Sharon Muir,⁴ Shahana Ferdousi,⁵ Samantha Moubarak,¹ Yalchin Oytam,¹ James Linden,¹ Louise Fisher¹

To cite: Correll P, Feyer A-M, Phan P-T, *et al.* Lumos: a statewide linkage programme in Australia integrating general practice data to guide system redesign. *Integrated Healthcare Journal* 2021;**3**:e000074. doi:10.1136/ihj-2021-000074

Received 15 January 2021
Revised 08 April 2021
Accepted 28 April 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Ministry of Health, NSW Health, St Leonards, New South Wales, Australia

²School of Computer Science, University of Technology Sydney, Sydney, New South Wales, Australia

³Hills Family General Practice, Castle Hill, New South Wales, Australia

⁴Riverstone Family Medical Practice, Riverstone, New South Wales, Australia

⁵Western Sydney Primary Health Network, Blacktown, New South Wales, Australia

Correspondence to

Patricia Correll;
patricia.correll@health.nsw.gov.au

ABSTRACT

Objective With ageing of the Australian population, more people are living longer and experiencing chronic or complex health conditions. The challenge is to have information that supports the integration of services across the continuum of settings and providers, to deliver person-centred, seamless, efficient and effective healthcare. However, in Australia, data are typically siloed within health settings, precluding a comprehensive view of patient journeys. Here, we describe the establishment of the Lumos programme—the first statewide linked data asset across primary care and other settings in Australia and evaluate its representativeness to the census population.

Methods and analysis Records extracted from general practices throughout New South Wales (NSW), Australia's most populous state, were linked to patient records from acute and other settings. Innovative privacy and security technologies were employed to facilitate ongoing and regular updates. The marginal demographic distributions of the Lumos cohort were compared with the NSW census population by calculating multiple measures of representation to evaluate its generalisability.

Results The first Lumos programme data extraction linked 1.3 million patients' general practice records to other NSW health system data. This represented 16% of the NSW population. The demographic distribution of patients in Lumos was >95% aligned to that of the NSW population in the calculated measures of representativeness.

Conclusion The Lumos programme delivers an enduring, regularly updated data resource, providing unique insights about statewide, cross-setting healthcare utilisation. General practice patients represented in the Lumos data asset are representative of the NSW population overall. Lumos data can reliably be used to identify at-risk regions and groups, to guide the planning and design of health services and to monitor their impact throughout NSW.

INTRODUCTION

Integrating care at a system level critically depends on understanding what patients need, where and when. Health information that is linked across services, settings and systems provides fundamental possibilities for informing better health decisions at the point of care. It can also support better

Significance of this study

What is already known about this subject?

- ▶ Integrating healthcare depends on understanding what patients need, where and when.
- ▶ To be effective in the delivery of coordinated, patient-centred health, data must also be integrated across healthcare settings.
- ▶ The integration of data is impeded in many health systems because of discrete system structures that inhibit information exchange.

What does this study add?

- ▶ Efforts in collaboration, technology and privacy management were key to the successful delivery of an enduring, cross-government, jurisdiction-wide linked data asset, in a highly siloed health system.
- ▶ Novel application of methodologies for quantifying the representativeness of a sample can be used effectively to compare the marginal distributions of demographic data from different data sources.
- ▶ The Lumos programme provided strongly representative distributions of selected key demographic variables, collectively across a range of statistical measures, compared with the New South Wales state population census data.

How might this impact on clinical practice or future developments?

- ▶ Lumos data can reliably provide insights about care that is delivered across the healthcare continuum, particularly for integrating healthcare.
- ▶ The Lumos data asset is uniquely placed to inform and improve the management of the health system in Australia and provides an exemplar for other contexts.
- ▶ The data can be used to inform the development of healthcare services and evaluate the efficacy of different care types to achieve better outcomes for patients.

understanding and decisions for managing population health. In many circumstances, the outcomes from one health sector's care can only be measured in another sector. Health information linked across services, settings



and systems, therefore provides an expanding evidence base that informs the redesign of health services. Existing research has highlighted the value of having primary care incorporated with other health-setting data as a basis for implementing integrated care interventions.¹⁻⁴

Australian structural arrangements have created obstacles for both the integration of care and for the smooth flow of information across the patient care continuum. Healthcare in Australia is provided through both private and public services and has separate Australian and state jurisdictional accountabilities. As a case in point, primary care is mostly delivered by private general practices that are funded partly or fully through Medicare, the Australian national health insurance scheme. Public hospital services are funded and administered by the individual state governments and private hospital services are largely funded through private insurers. Furthermore, Primary Health Networks (PHNs) are federally funded regional commissioners of care operated by not-for-profit companies and make decisions that are independent of government regarding which primary healthcare services or interventions should be provided and who should provide them. PHNs work closely with primary health providers to improve performance and facilitate change. There are 31 PHNs distributed across Australia including 10 in New South Wales (NSW). These discrete system structures mean that information exchange is complex. While there is broad-based agreement regarding the value of detailed, timely and comprehensive data to achieve effective and integrated healthcare in Australia,⁵ access remains complicated because health data are collected by different arms of government as well as by private organisations. Whole-of-system information is not available to any one steward. This separation of data means the view of patient journeys—the care that patients receive and the outcomes that they experience—is fragmented; to individual consumers, their families, health providers and health researchers.^{6,7}

Lack of comprehensive data has been particularly challenging in understanding care across primary and acute settings in Australia. An estimated 80% of Australians visit a general practitioner at least once in a year.⁸ General practices are usually the first point of contact with the health system and they often fill the gaps between episodes of other healthcare. As a result, general practices hold invaluable information about individual patient health and care,^{9,10} often from the start of a patient's health journey. Our ability to understand how primary care might relate to the performance of our health system, and the implications for improving delivery has remained limited.¹¹ Addressing the lack of access to general practice record data is challenging in the Australian context. There are considerations about patient privacy, private business practices, diversity of software platforms and interoperability between general practice systems and the rest of the health system.¹²⁻¹⁴ Currently, there are no agreed or common standards for practice management software. Fortunately, in the

contemporary health landscape, these barriers are no longer insurmountable.

Record linkage, an established tool for bringing together data from multiple sources about the same individuals, is an obvious candidate for solving the lack of access to general practice data to understand healthcare across settings.^{15,16} The use of data linkage is growing in Australia and elsewhere,^{17,18} and offers substantial benefits in a fragmented healthcare system such as Australia's.¹⁹ Historically, access to general practice patient records for linkage has been very restricted in Australia. Linkage incorporating data from Medicare claims items has gone some way towards this ambition,²⁰ however, this information largely lacks clinical details about care events. Much of that clinical information resides within the clinical management systems of general practices.

NSW is Australia's most populous state with an estimated resident population exceeding 8 million people.²¹ As elsewhere, the NSW health system is under continual pressure to evolve in response to emerging population health needs, changing patterns of disease, changing health service priorities, budgetary constraints and evolving evidence. In recent years, NSW Health strategic planning has focused on better integrating state services with primary care and 'Delivering Truly Integrated Care' is one of three strategic directions of the current State Health Plan.²² Information that maps across the healthcare system is needed to support this vision.

Between 2016 and 2019, NSW Ministry of Health collaborated with NSW PHNs, undertaking the Lumos pilot project to explore the feasibility and value of extracting and linking data from general practice records to other health system records. The Lumos pilot project demonstrated the feasibility of extracting patient records from general practices and linking to other health system records, and the potential public benefits of the data linkage to underpin integrated, efficient and effective care. The current Lumos programme has scaled-up the pilot project to implement a programme of work throughout NSW.

To our knowledge, this is the first Australian programme for the creation of an enduring, statewide and routinely updated data asset incorporating general practice data with broad-ranging data linkage across the healthcare system, which can be used to inform and improve the management of an integrated health system in Australia. Here, we describe establishment of the linked data asset and present an assessment of the generalisability of Lumos data for informing integrated healthcare and monitoring patient outcomes in NSW.

MATERIALS AND METHODS

Design and setting

The Lumos programme creates a retrospective cohort comprising all patients who attended a participating general practice at least once since January 2010 or the earliest data available after that date. Twice each year,

extracts are taken from these general practice records and linked to other data collections in NSW Health over the same period. The linked data collections comprise hospitalisations, emergency department presentations, cancer registrations, integrated care, mental health services, death registrations and causes.

The Lumos programme builds on a pilot project (executed from 2015 to 2019) and is being implemented over a 3-year scale-up period from mid-2019 to mid-2022. It operates in all 10 PHN regions in NSW.

Patient and public involvement

The Lumos programme operates with ethical approval and has been granted a waiver of individual patient consent for secondary use of individual health information for the purposes of health services management, including funding, planning and evaluation. This approval permits linkage of patient health records in the NSW Health Centre for Health Record Linkage (CHeReL) after the ethics committee's assessment that there is high likelihood of public benefit in the Lumos programme with minimal risks to individual privacy. Patients have not been involved in the design, processes, or dissemination of results however, the waiver of consent does not preclude individual practices from notifying their patients of the Lumos programme, and they are able to exclude an individual's data from extraction if the patient asks to do so without a requirement to notify the Lumos programme. The ethics committee determination considered that the Lumos programme only accesses previously collected information, does not use directly identifying particulars (such as names of patients) for data linkage and presents no additional burden to patients. The de-identified linked data are made available to suitably qualified health system personnel under controlled access conditions to use for purposes as prescribed in the ethically approved protocol.

Recruitment and participation

In partnership with the Lumos programme, PHNs seek voluntary informed consent for data collection from general practices in their regions. PHNs are responsible for liaising with general practices, delivering written programme information and answering questions. Practices may have multiple opportunities to participate before they enrol. General practices provide written consent to enrol and are registered with the Lumos programme as participants. Once a practice has consented, it can continue to participate in subsequent linkages without further enrolment, or it may withdraw at any time. To date, there have been no withdrawals.

The Lumos programme is currently undertaking a 3-year scale-up process, which will progressively recruit practices so that by the end of this phase, over 500 practices will be regularly contributing data. The programme will be subject to an evaluation to determine its value for continued expansion. This first phase of scale up will result in approximately 15% of NSW general practices

participating in the Lumos programme. Experience from the preceding pilot project suggests this will deliver generalisable patient information because it will likely result in approximately 50% of the NSW population being included (around 4 million people). This is because general practices participating in Lumos tend to be those with larger patient numbers. Also, in Australia, patients can engage with multiple practices. So there will be patients included in Lumos data that also attended practices that were not participating in the Lumos programme.

Data management

Recruitment, data extraction and linkage occur twice per annum. With each extraction-linkage cycle, the data are completely refreshed. That is, all available records are relinked and re-processed. It is not possible to connect patients across different linkage episodes as the project person numbers change with each linkage. There are distinct advantages in this approach. First, linkage quality may improve over time as more information becomes available. Second, practices or individuals who withdraw will be removed from the Lumos data asset as soon as data are refreshed. The superseded data are retained and destroyed in accordance with the approved data retention protocol.

Data extraction

Once recruited, extraction and transfer of data from general practices are fully automated. The Lumos programme works with commercial general practice software providers to deploy data extraction software into the clinical information systems of recruited general practices. The extraction process, developed in conjunction with Curtin University in Western Australia, replaces patient details such as name and address, at source with encoded information, using Bloom filters.²³ The software automates regular encoding, extraction and transfer of data. Linkage files and content files are transferred separately via secure file transfer. General practice data extracts contain a range of health-related data from patient general practice records including administrative, demographic, clinical, lifestyle and physiological measures. These records are linked to health, service use, demographic and mortality data in other NSW health data collections.

Data linkage

Data linkage is performed by NSW Ministry of Health's dedicated data linkage facility, the CHeReL, which has access to the latest available combined statewide datasets for all data collections included in the Lumos programme. The encoded general practice data is probabilistically linked to the encoded data of other data sources in a process referred to as privacy preserving record linkage (PPRL). Current research in this field has demonstrated that PPRL using Bloom filters is effective at delivering valid information with greater privacy assurances and minimal loss of linkage accuracy.²⁴

The CHeReL implements best-practice linkage²⁵ and follows the separation principle,²⁶ whereby linkage information is stored and accessed separately from content information. The CHeReL has separate data linkage and data integration units, and therefore also assembles the linked data asset for delivery to the Lumos programme. The final linked data asset contains health-related information only. Records pertaining to the same individual are distinguished by a project person number which is unique to both the patient and the linkage episode. The process of de-identification is one-way such that patient identities cannot be re-assigned to the linked data asset, even by the source practice. Any residual risk of re-identification is very low and well managed.

Data storage and access

NSW Ministry of Health has developed a highly secure analysis environment—the Secure Analytics Primary Health Environment (SAPHE) as a part of the Lumos programme. The SAPHE is a single, highly secure cloud-based platform that allows secure access to authorised Lumos collaborators. It comprises a multiuser business intelligence and data analytics platform and includes multilevel access controls for users with different needs, authorisations and security clearances.

Statistical reporting

Summary statistics were generated from the first scale-up data extraction to characterise the linked Lumos data asset including patient demographics, service use and health-related risk factors. Patient residential locality data were used to generate socioeconomic status using the Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)²⁷ and remoteness area using the Accessibility

and Remoteness Index of Australia (ARIA),²⁸ which are both derived from the Australian Census.

To address the principal aim of this analysis, the distributions of four demographic statistics: (1) age, (2) sex, (3) IRSAD and (4) ARIA were compared with the overall NSW population 2020 projections from the most recent census in 2016. We focused on population demographic and geographical distributions to compare official census estimates that would have greater independence, greater representative and less bias than estimates derived from samples. We assessed the similarity of Lumos data to census data by computing the marginal probability distributions of these four demographic indices for each dataset (Lumos and the NSW census). We compared each marginal distribution by computing a range of measures of representativeness.

RESULTS

The first scale-up data extraction and linkage was completed in May 2020. There were 1 484 314 general practice patient records extracted from the 156 participating NSW general practices who had attended those practices one or more times in the 5-year period between May 2015 and May 2020. Of these, 181 959 (12%) had attended more than one of the included practices, leaving 1 302 355 individual general practice patients, or approximately 16% of the NSW population. Their health-related journeys were linked to the most recent available records in NSW hospitals, emergency departments, mental health services, ambulance records, non-admitted patient services, integrated care enrolments, cancer registrations and mortality records.

Table 1 Lumos data linkage results and patient demographic profile, May 2020

Data source	General practice records	Public hospitals	Private hospitals	Emergency department	Mental health ambulatory services	Non-admitted patient services	NSW ambulance records	Integrated care	Cancer registry	Death registry	Causes of death
Time period of data availability	Mid-May 2015–mid-May 2020	January 2010–December 2019	January 2010–30 June 2019	January 2010–December 2019	January 2010–December 2019	July 2015–December 2019	January 2010–September 2019	January 2014–mid-February 2020	January 2010–December 2016	January 2010–December 2019	January 2010–December 2017
Number of linked records	1 484 314	2 281 894	1 475 432	4 125 094	5 108 582	8 842 592	601 189	9309	28 395	24 141	11 208
Number of linked patients	1 302 355	638 602	393 589	871 780	93 285	562 632	248 676	9309	27 081	24 141	11 208
Per cent of linked patients	Reference	49%	30%	67%	7%	43%	19%	1%	2%	2%	1%
Average records per person	1.1	3.6	3.7	4.7	54.8	15.7	2.4				
Median age (years)	37	39	50	37	36	38	50	67	71	82	82
Gender (%)											
Male	46.58	44.34	42.94	48.32	45.01	43.24	47.04	50.47	52.67	52	52
Female	53.27	55.66	57.06	51.68	54.99	56.76	52.96	49.53	47.33	48	48

NSW, New South Wales.

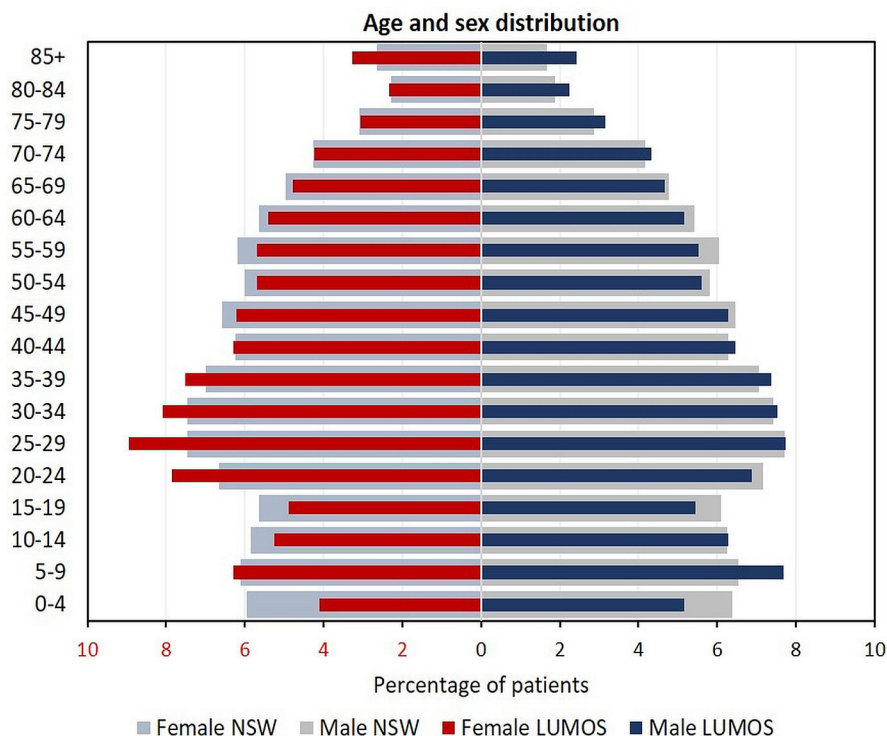


Figure 1 Age and sex distribution, Lumos patients May 2020 and New South Wales (NSW) population 2020.

Linkage results are shown in [table 1](#). Of the included general practice patients, 49% had been admitted to a public hospital, 67% had an emergency department presentation and 2% died between 1 January 2010 and 31 December 2019. In comparison with the overall NSW population, Lumos participants had similar median age (NSW 38 years, Lumos 37 years, data not shown) and slightly more females (NSW 50.7%, Lumos 53.3%, data not shown). The distribution of age and sex in Lumos data compared with that of the NSW population is displayed in [figure 1](#). This suggests there were slightly fewer young children, slightly more women aged 20–39 years and slightly more very elderly (age 85+ years) in Lumos data. The regional distribution of Lumos patients in comparison to NSW shows slightly higher representation in inner

and outer regional areas ([figure 2](#)) and slightly less individuals lived in the most socioeconomically advantaged areas ([figure 3](#)).

Overall, the seven measures of representativeness calculated in this analysis consistently indicated that the distribution of the first Lumos programme scale-up data extraction was highly representative of the NSW census population for these demographic indices ([table 2](#)). Of our reported measures, the most interpretable are:

- Jensen-Shannon (J-S) divergence, which can be interpreted as the probability of a misclassification error, where 0 implies the distributions are identical and 1 is the maximum difference.²⁹ Our results show the probability bound on misclassification error is <1%.

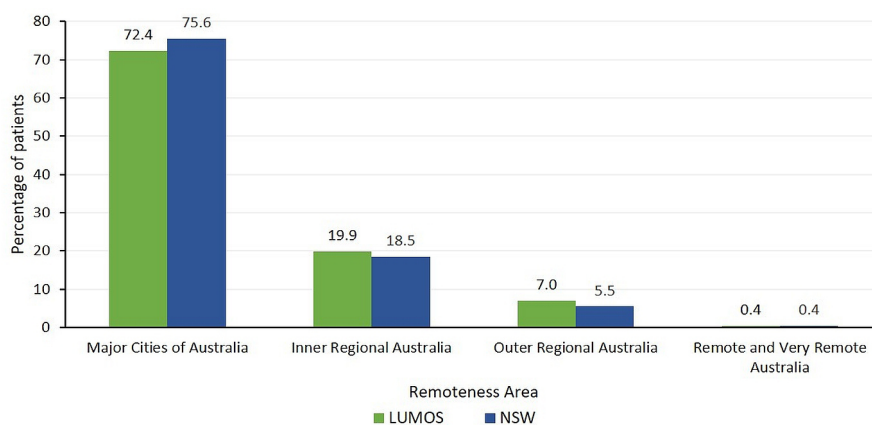


Figure 2 Remoteness area of residence (Accessibility and Remoteness Index of Australia (ARIA)), Lumos patients May 2020 and New South Wales (NSW) population 2020.

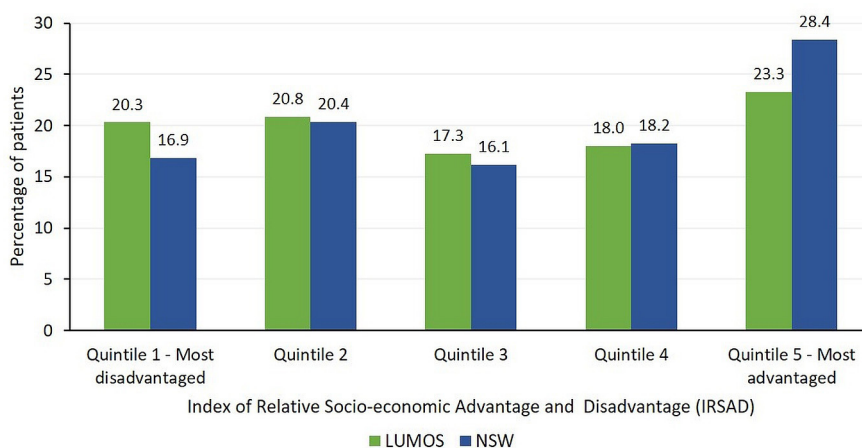


Figure 3 Socioeconomic status (Index of Relative Socioeconomic Advantage and Disadvantage quintiles (IRSAD)), Lumos patients May 2020 and New South Wales (NSW) population 2020.

- ▶ Bhattacharyya distance, which is an approximate measurement of the amount of overlap between two statistical samples, where 0 implies no overlap and 1 implies they are identical.³⁰ Our results show over 99% overlap.
- ▶ Histogram intersection, which measures the similarity of two histograms, where 0 implies no intersection and 1 implies they are identical. Our results show over 95% intersection.³¹

The total variation³² and Hellinger distance coefficient^{33 34} are related to the Bhattacharyya distance, reported here for completeness. We also report on the Kullback-Leibler divergence as it is a measure of the information lost using Lumos data as a surrogate for census data.³⁵

Finally included is the Kolmogorov-Smirnov (K-S) test results. The K-S statistic measures the difference between two distributions.³⁶ This test detected that the distributions were statistically different, but this metric is of limited use because it does not indicate the meaningfulness of the difference.

DISCUSSION

The Lumos programme implements data systems and infrastructure where the records of patients who have attended participating general practices are linked to data collections in the NSW health system. It directly aims to address the need for better evidence to guide strategic policy directions.^{5 22} A great strength of the programme is its scale and reach. Currently, Lumos data include information about approximately 16% of the NSW population, and is projected to increase to approximately 50% by mid-2022. General practice patients represented in the Lumos data asset are representative of the NSW population overall, although with slightly more representation in both regional and less socioeconomically advantaged areas. We assessed representativeness of Lumos data to the NSW census population using multiple approaches to compare the marginal distributions of demographic data from each dataset. To our knowledge, there are no broadly agreed methodologies or measures for quantifying the representativeness of a sample, however, taken

Table 2 Representativeness of marginal distributions of four demographic indices from Lumos data to the NSW census 2016 data

Measure of representativeness	5-year age groups—female	5-year age groups—male	ARIA (remoteness)	IRSAD quintiles (socioeconomic status)
J-S divergence (binary log)	0.00284	0.00114	0.00104	0.00276
Bhattacharyya coefficient	0.998	0.999	0.999	0.998
Histogram intersection	0.952	0.971	0.971	0.951
Hellinger distance	0.0444	0.0282	0.0269	0.0438
Total variation distance	0.0481	0.0291	0.0285	0.0488
K-L divergence (natural log)	0.00780	0.00317	0.00297	0.00763
K-S statistic	0.0266	0.0112	0.0285	0.0488
K-S test, p value	<2.22e-16	<2.22e-16	<2.22e-16	<2.22e-16

ARIA, Accessibility and Remoteness Index of Australia; IRSAD, Index of Relative Socioeconomic Advantage and Disadvantage quintiles; J-S, Jensen-Shannon; K-L, Kullback-Leibler; K-S, Kolmogorov-Smirnov; NSW, New South Wales.

collectively, these measures show that at least the marginal distributions of selected key demographic variables are representative of the NSW population census data.

For the first time in Australia, Lumos provides general practice data and information on the journey of a representative primary care patient cohort across the tertiary health sector. The Lumos programme will enable system-wide strategic planning and planning at the regional and local level down to the individual general practice. By triggering reflection on a practice population and their comorbidities, it will also help to drive improvement in the quality of general practice data and health records.

Long-standing challenges have hindered efforts to access primary care data to understand the whole patient journey and inform healthcare in Australia. The success of the Lumos programme reflects that three leading (and interrelated) challenges have been actively targeted: building collaborations, developing technological solutions for access to data and managing privacy and security.

First, fragmentation of health service delivery between the federal and state governments in Australia does not naturally motivate collaborative endeavours. Strategic imperatives to develop greater integration of healthcare services have been a catalyst for bringing the private general practice and state public health systems together. A critical element contributing to the success of the Lumos programme to date is attributable to collaboration and common goals. Key to this has been harnessing vested interests in recognition of the potential opportunities of the Lumos programme for delivering information that can improve patient care and outcomes. Focussing on the benefits to each stakeholder, as well as identifying shared purposes (good patient outcomes), is a critical element for the Lumos programme's continued success.

The development of cross-sectoral relationships is inexorably linked to the second key challenge that the Lumos programme has directly targeted: access and the technology needed to achieve it. NSW Health data assets are collected and held under enabling legislation.³⁷ Access is made available to people outside of NSW Health by authorised release of copies of data. However, these copies lose their currency as the source data continues to grow and to be maintained. The Lumos data asset is a regularly updated aggregate of NSW Health System data and general practice record data. This is borne of an enduring partnership between all data custodians. The challenge is to provide access to the current data for all users, without a proliferation of uncontrolled, copies. The response to this challenge has focused on investment in the development and implementation of technology to offer safe and secure data access to all collaborators through the implementation of a controlled, cloud-based environment: the SAPHE.

Another technology challenge has been the lack of interoperability of general practice clinical information systems resulting from the range of software technologies used by general practices for managing patient records. The Lumos programme must necessarily be independent

of vendors and employ highly automated processes to manage the burden of its activities. Working with a range of commercial software vendors to create and transfer general practice electronic patient record extracts remains an ongoing endeavour.

Finally, privacy is a major challenge that the Lumos protocol has addressed through defined processes and technological innovation. The Lumos programme has been granted a waiver of individual patient consent for the disclosure and use of patient records from participating general practices via ethical review. This provides the legal provision that permits general practices to disclose patient data to the Lumos programme based on it having significant public benefit for the management of health services at minimal risk to patient privacy. The Lumos programme does not have a mandate to require or enforce patient agreements with general practices to receive general practice records. This approach has further been subjected to an independent privacy impact assessment which endorsed its approaches.³⁸ A contributing factor has been that the Lumos programme demonstrates the large-scale application of PPRL with Bloom filters, using previously validated methodologies to deliver high quality linkage.³⁹ The use of PPRL techniques, which enables linkage of data without directly identifying patient details (such as name and address) is an important design feature that reduces potential privacy risks, supports practice engagement and bolsters community acceptability.

In Australia, patients are free to go to any available general practitioner. This differs to other places, for example, in the UK and New Zealand, where, in the mechanism for providing universal healthcare, patients typically enrol with a general practice on an ongoing basis.⁴⁰⁻⁴¹ These structural arrangements elsewhere allow more complete capture of delivery of care from general practices. In Australia, there is no assurance that data from any one general practice captures the entire record of an individual's general practice interactions. Patients may visit multiple general practices, some of whom may contribute data to the Lumos programme, and others not. However, better integration and continuity of care, through a model focused on enrolment of Australians 70 years and older with their general practice, has been recently foreshadowed as part of the Commonwealth Government's Long Term National Health Plan.⁴² If this or other patient enrolment/registration strategies are successful, it would reduce incompleteness in general practice record coverage.

The practices represented in the Lumos programme may also not represent the full range of primary care offered and/or received in the community; the participating practices may fundamentally differ to those not participating. Lumos data may therefore lack some aspects of general practice experience. As the Lumos programme expands and includes a greater proportion of NSW general practices, these limitations stand to diminish, particularly as the Lumos programme will predominantly engage with



large-scale practices. However, in the absence of universal patient enrolment with a designated general practice, as is the case in New Zealand and the UK, these limitations are unlikely to ever be fully eliminated.

The underlying purposes for which general practices maintain records primarily relate to managing their private practices and business processes. This yields data of varying quality for the purposes of integrating information about patient journeys for service development and improvement. Recent implementation of quality improvements payment incentives for general practice care may support improvements in data quality as well as care.⁴³ Addressing the limitations in general practice data quality and standards is a necessary area for ongoing development that will support the sustainability of the Lumos programme and should align to existing efforts by others in this field.^{44 45}

CONCLUSION

Compared with other places, the UK and New Zealand for example, Australia has lagged in having consistent and accessible patient care information from general practices that can support strategy, policy and reform. The imperative in Australia to create the data-based information needed to support health system reform in Australia is now well established. The focus has shifted from justifying ‘why’ create such data, to, ‘how’ to create it. Effort now concentrates on the processes needed to facilitate building secure, safe data assets that can deliver the much-needed insights for healthcare system redesign. The Lumos programme demonstrates a robust approach that is feasible within the existing structures of the Australian health system.

Acknowledgements The following individuals supported the development of the Lumos protocol underlying this article by providing their expert input and advice: James Boyd, Jean-Frederic Levesque, Sharon Smith, Ian Corless, Ray Messom.

Contributors All authors have approved this version of the article to be published and agreed to be accountable for all aspects of the work. In addition, the authors made the following individual contributions: PC conceived and designed the study, interpreted the findings and drafted the manuscript and revisions; A-MF conceived and designed the study, interpreted the findings and drafted the manuscript and revisions; TP analysed and interpreted the data and drafted the manuscript; BD analysed and interpreted the data and drafted the manuscript; WJ advised on the study design, acquired data and drafted the manuscript; KI advised on the study design, acquired data and revised the manuscript; AP acquired data and revised the manuscript; SM advised on the study design, acquired data and revised the manuscript; SF acquired data and revised the manuscript; SM advised on the study design and revised the manuscript; YO analysed data and revised the manuscript; JL advised on the study design and revised the manuscript; LF conceived and designed the study, interpreted the findings and revised the manuscript.

Funding The Lumos programme has been funded by the Commonwealth Government under the Health Innovation Fund.

Competing interests WJ holds a position on the editorial board of *BMJ Integrated Healthcare Journal*. None other declared.

Patient consent for publication Not required.

Ethics approval The Lumos programme operates with ethical approval from the NSW Population and Health Services Research Ethics Committee (approval ID 2019/ETH00660). The ethics committee approved a waiver of individual patient consent

for secondary use of individual health information for the purposes of health services management.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. NSW Ministry of Health has developed a highly secure data repository and analysis environment; the Secure Analytics Primary Health Environment (SAPHE) as a part of the Lumos programme. The SAPHE is a single, highly secure platform that allows authorised users to securely access and analyse data from the Lumos programme. Approved personnel from NSW Primary Health Networks and NSW Health can access Lumos data through the SAPHE, which provides the necessary access for all collaborators to successfully benefit from participating in the Lumos programme while adhering to strict privacy and security controls for the management of personal health information.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- 1 Khanna S, Rolls DA, Boyle J, *et al*. A risk stratification tool for hospitalisation in Australia using primary care data. *Sci Rep* 2019;9:5011.
- 2 Johnson TL, Kaldor J, Falster MO, *et al*. Predictive risk modelling under different data access scenarios: who is identified as high risk and for how long? *BMJ Open* 2018;8:e018909.
- 3 Pearce C, McLeod A, Rinehart N, *et al*. Polar diversion: using general practice data to calculate risk of emergency department presentation at the time of consultation. *Appl Clin Inform* 2019;10:151–7.
- 4 Dennis S, Taggart J, Yu H, *et al*. Linking observational data from general practice, hospital admissions and diabetes clinic databases: can it be used to predict hospital admission? *BMC Health Serv Res* 2019;19:526.
- 5 Council of Australian Governments. Heads of agreement on public hospital funding and health reform. Canberra COAG; 2018. <https://www.coag.gov.au/about-coag/agreements/heads-agreement-between-commonwealth-and-states-and-territories-public-0> [Accessed Sep 2020].
- 6 Srinivasan U, Rao S, Ramachandran D. Flying blind: Australian consumers and digital health. Sydney Australian Consumers and Digital Health. Australian Health Data Series Volume 1; 2016. <https://flyingblind.cmcrc.com/files/files/Flying-Blind-Australian-Consumers-and-Digital-Health.pdf> [Accessed Jul 2020].
- 7 Srinivasan U, Ramachandran D, Quilty C. Flying blind: Australian researchers and digital health. Australian health data series volume 2. Sydney Centre DHCR; 2018. https://flyingblind.cmcrc.com/files/cmcrc_flying_blind_vol_2_web4.pdf [Accessed Jul 2020].
- 8 Britt H, Miller G, Henderson J. General practice activity in Australia 2013–14: general practice series. Sydney Press SU; 2014. http://ses.library.usyd.edu.au/bitstream/2123/11882/4/9781743324226_ONLINE.pdf [Accessed Sep 2020].
- 9 Duckett S, Swerissen H. Building better foundations for primary care. Melbourne Grattan Institute; 2017. <https://grattan.edu.au/wp-content/uploads/2017/04/Building-better-foundations-for-primary-care.pdf> [Accessed Sep 2020].
- 10 The Royal Australian College of General Practitioners. Standards for clinical practice, 2020. Available: <https://www.racgp.org.au/FSEDEV/media/documents/Running%20a%20practice/Practice%20standards/5th%20edition/Standards-for-general-practice-5th-edition.pdf> [Accessed Sep 2020].
- 11 Australian Institute of Health and Welfare. Review and evaluation of Australian information about primary health care: a focus on general practice. CAT. No. HWI 103. Canberra AIHW; 2008. <https://www.aihw.gov.au/getmedia/aec476dc-21da-48a4-b33a-e0b7d7c71f8c/reaiaphc-c00.pdf.aspx> [Accessed Sep 2020].
- 12 Palamuthusingam D, Johnson DW, Hawley C, *et al*. Health data linkage research in Australia remains challenging. *Intern Med J* 2019;49:539–44.
- 13 Boyle DIR. Middleware supporting next generation data analytics in Australia. *Stud Health Technol Inform* 2015;216:1019.
- 14 The Royal Australian College of General Practitioners. Minimum requirements for general practice clinical information systems to improve usability. East Melbourne, VIC RACGP Practice Technology and Management; 2019. <https://www.racgp.org.au/FSEDEV/media/>

- documents/Running%20a%20practice/Support%20and%20tools/Minimum-requirements-for-general-practice-CIS.pdf [Accessed Sep 2020].
- 15 Bohensky MA, Jolley D, Sundararajan V, *et al.* Data linkage: a powerful research tool with potential problems. *BMC Health Serv Res* 2010;10:346.
 - 16 Centre for Health Record Linkage. Centre for health record linkage website. Sydney NSW Ministry of Health; 2020. <https://www.cherel.org.au/our-services> [Accessed Sep 2020].
 - 17 Emery J, Boyle D. Data linkage. *Aust Fam Physician* 2017;46:615–9.
 - 18 Waller P, Cassell JA, Saunders MH, *et al.* Governance and oversight of researcher access to electronic health data: the role of the independent scientific Advisory Committee for MHRA database research, 2006–2015. *J R Coll Physicians Edinb* 2017;47:24–9.
 - 19 Boyd JH, Ferrante AM, O'Keefe CM, *et al.* Data linkage infrastructure for cross-jurisdictional health-related research in Australia. *BMC Health Serv Res* 2012;12:480.
 - 20 Briffa TG, Jorm L, Jackson RT, *et al.* Nationally linked data to improve health services and policy. *Med J Aust* 2019;211:397–8.
 - 21 Australian Bureau of Statistics. National, state and territory population March 2020. Cat 3101.0 Canberra: ABS, 2020. Available: <https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release#key-statistics> [Accessed Sep 2020].
 - 22 NSW Ministry of Health. NSW state health plan: towards 2021. Sydney NSW Ministry of Health; 2014. <https://www.health.nsw.gov.au/statehealthplan/Publications/NSW-State-Health-Plan-Towards-2021.pdf> [Accessed Sep 2020].
 - 23 Schnell R, Bachteler T, Reiher J. Privacy-preserving record linkage using Bloom filters. *BMC Med Inform Decis Mak* 2009;9:41.
 - 24 Brown AP, Ferrante AM, Randall SM, *et al.* Ensuring privacy when integrating patient-based datasets: new methods and developments in record linkage. *Front Public Health* 2017;5:34.
 - 25 Irvine K, Hall R, Taylor L. Centre for health record linkage: expanding access to linked population data for NSW and the act, Australia. *International Journal of Population Data Science* 2019;4.
 - 26 Kelman CW, Bass AJ, Holman CDJ. Research use of linked health data - a best practice protocol. *Aust N Z J Public Health* 2002;26:251–5.
 - 27 Australian Bureau of Statistics. Socio-Economic indexes for areas (SEIFA). Technical paper Cat 2033.0.55.001. Canberra ABS; 2018. [https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756E3DBEFA869EFC258259000BA746/\\$File/SEIFA%202016%20Technical%20Paper.pdf](https://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/756E3DBEFA869EFC258259000BA746/$File/SEIFA%202016%20Technical%20Paper.pdf) [Accessed Oct 2020].
 - 28 Australian Bureau of Statistics. Australian Statistical Geography Standard (ASGS): Volume 5 - Remoteness Structure, July 2016. Cat 1270.0.55.005. Canberra ABS; 2018. <https://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.005> [Accessed Oct 2020].
 - 29 Lin J. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory* 1991;37:145–51.
 - 30 Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society* 1943;35:99–109.
 - 31 Swain MJ, Ballard DH. Color indexing. *Int J Comput Vis* 1991;7:11–32.
 - 32 Rachev S. *Probability Metrics and the Stability of Stochastic Models (Wiley Series in Probability and Statistics - Applied Probability and Statistics Section)*. 512. 1st Edition. Wiley-Blackwell, 1991.
 - 33 Hellinger E. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik* 1909;136:210–71.
 - 34 Nikulin M. Hellinger distance. In: *Encyclopedia of mathematics*. EMS Press, 2001.
 - 35 Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat* 1951;22:79–86.
 - 36 Massey FJ. The Kolmogorov-Smirnov test for goodness of fit. *J Am Stat Assoc* 1951;46:68–78.
 - 37 Health Administration Act 135 NSW 1982.
 - 38 Johnston A. Privacy impact assessment of the Lumos program in report delivered to NSW Ministry of health client in confidence. Sydney Salinger Privacy; 2020.
 - 39 Randall SM, Ferrante AM, Boyd JH, *et al.* Privacy-preserving record linkage on large real world datasets. *J Biomed Inform* 2014;50:205–12.
 - 40 National Health Service. How to register with a doctor (GP). England NHS England Publications Gateway; 2020. <https://www.nhs.uk/NHSEngland/AboutNHSservices/doctors/Documents/how-to-register-with-a-gp-leaflet.pdf> [Accessed 04 Aug 2020].
 - 41 New Zealand Ministry of Health. Enrolment in a primary health organisation [web page]. New Zealand: New Zealand Government; 2020. Available from. Available: <https://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations/enrolment-primary-health-organisation> [Accessed Sep 2020].
 - 42 Department of Health. Australia's Long Term National Health Plan, 2019. Available: https://www.health.gov.au/sites/default/files/australia-s-long-term-national-health-plan_0.pdf [Accessed Sep 2020].
 - 43 Australian Government Department of Health. Practice incentives program quality improvement incentive guidelines, 2020. Available: [https://www1.health.gov.au/internet/main/publishing.nsf/Content/46506AF50A4824B6CA25848600113FFF/\\$File/Practice-Incentives-Program-Quality-Improvement-Incentive-Guidelines.pdf](https://www1.health.gov.au/internet/main/publishing.nsf/Content/46506AF50A4824B6CA25848600113FFF/$File/Practice-Incentives-Program-Quality-Improvement-Incentive-Guidelines.pdf)
 - 44 Taggart J, Liaw S-T, Yu H. Structured data quality reports to improve EHR data quality. *Int J Med Inform* 2015;84:1094–8.
 - 45 Liyanage H, Liaw S-T, Jonnagaddala J, *et al.* Common data models (CDMs) to enhance international big data analytics: a diabetes use case to compare three CDMs. *Stud Health Technol Inform* 2018;255:60–4.