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,3 Shahana Ferdousi,5 Samantha Moubarak,1 Yalchin Oytam,1 James Linden,1 Louise Fisher1

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 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 probabilistically linked data. The software automates regular encoding, extraction and transfer of data. Data 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 linkage and follows the separation principle, 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) and remoteness area using the Accessibility and Remoteness Index of Australia (ARIA), 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 2016 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 1484314 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, 181959 (12%) had attended more than one of the included practices, leaving 1302355 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

<table>
<thead>
<tr>
<th>Data source</th>
<th>General practice records</th>
<th>Public hospitals</th>
<th>Private hospitals</th>
<th>Emergency department</th>
<th>Mental health ambulatory services</th>
<th>Non-admitted patient services</th>
<th>NSW ambulance records</th>
<th>Integrated care</th>
<th>Cancer registry</th>
<th>Death registry</th>
<th>Causes of death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time period of data availability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-May 2015–mid-May 2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of linked records</td>
<td>1484314</td>
<td>2281894</td>
<td>1475432</td>
<td>4125094</td>
<td>5108582</td>
<td>8842592</td>
<td>601189</td>
<td>9309</td>
<td>28395</td>
<td>24141</td>
<td>11208</td>
</tr>
<tr>
<td>Number of linked patients</td>
<td>1302355</td>
<td>638602</td>
<td>393589</td>
<td>871780</td>
<td>93285</td>
<td>562632</td>
<td>248676</td>
<td>9309</td>
<td>27081</td>
<td>24141</td>
<td>11208</td>
</tr>
<tr>
<td>Per cent of linked patients</td>
<td>Reference</td>
<td>49%</td>
<td>30%</td>
<td>67%</td>
<td>7%</td>
<td>43%</td>
<td>19%</td>
<td>1%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Average records per person</td>
<td>1.1</td>
<td>3.6</td>
<td>3.7</td>
<td>4.7</td>
<td>54.8</td>
<td>15.7</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median age (years)</td>
<td>37</td>
<td>39</td>
<td>50</td>
<td>37</td>
<td>36</td>
<td>38</td>
<td>50</td>
<td>67</td>
<td>71</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>Gender (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>46.58</td>
<td>44.34</td>
<td>42.94</td>
<td>48.32</td>
<td>45.01</td>
<td>43.24</td>
<td>47.04</td>
<td>50.47</td>
<td>52.67</td>
<td>52</td>
<td>52</td>
</tr>
<tr>
<td>Female</td>
<td>53.27</td>
<td>55.66</td>
<td>57.06</td>
<td>51.68</td>
<td>54.99</td>
<td>56.76</td>
<td>52.96</td>
<td>49.53</td>
<td>47.33</td>
<td>48</td>
<td>48</td>
</tr>
</tbody>
</table>

NSW, New South Wales.
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. Our results show the probability bound on misclassification error is <1%.
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. 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.

The total variation and Hellinger distance coefficient 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.

Finally included is the Kolmogorov-Smirnov (K-S) test results. The K-S statistic measures the difference between two distributions. 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. 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...
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. Focusing 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 inextricably linked to the second key challenge that the Lumos programme has directly targeted: access and use 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.36 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 may support improvements in data quality as well as care. 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.

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-Frédéric Levesque, Sharon Smith, Ian Corless, Ray Misson.

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; AMF 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; YO analysed data and revised the manuscript; SF acquired data and revised the manuscript; SM advised on the study design, acquired data and drafted the manuscript; KI advised on the study design and revised the manuscript; YO analysed data and revised the manuscript; LS advised on the study design and revised the manuscript; LJ 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

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.

Appendix

Provenance and peer review
Not commissioned; externally peer reviewed.

Data availability statement
No data are available.

REFERENCES


37 Health Administration Act 135 NSW 1982.


