Communicating statistical outputs through health maps

Prepared for the National Health Performance Authority

June 2016

Suggested citation

Roberts JL, Cramb SM, Baade PD, Mengersen KL, 2016. *Communicating Statistical Outputs Through Health Maps*. Brisbane: Cancer Council Queensland and Queensland University of Technology (QUT).

Author affiliations

ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology (QUT): Jessie Roberts, Susanna Cramb and Kerrie Mengersen.

Viertel Cancer Research Centre, Cancer Council Queensland: Susanna Cramb and Peter Baade.

Executive Summary

Disease mapping is a powerful tool for exploring spatial variation of health and disease outcomes. The methods behind these maps can be highly sophisticated, and the insights derived from these analyses can be valuable to policy-makers and other decision makers. However, the impact of these maps to create changes in real world health outcomes is limited by our ability to effectively and efficiently communicate to key decision makers.

The following report aims to be a resource that will guide the visual design of the National Cancer Atlas as well as the overarching communication strategy that cancer maps will sit within. The information provided aims to ensure that communication and design decisions made in the development of the National Cancer Atlas are well designed and comprehensively informed.

This report explores the visualisation platforms used for disease mapping, both current and emerging. The importance of uncertainty communication and the current approaches to uncertainty visualisation are discussed. The report concludes with a report of a communication design workshop that was conducted with the main stakeholders of the National Cancer Atlas project.

This report is a summary of the research conducted to date, and will continue to be developed in the future. We hope it will be an informative resource as we move forward with the design of the National Cancer Atlas.

Contents

List of Tables

List of Figures

List of Abbreviations

CI	Confidence interval
CCQ	Cancer Council Queensland
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CSS	Cascading Style Sheets
DOM	Document Object Model
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System
HTML	HyperText Markup Language
NCI	National Cancer Institute
NHPA	National Health Performance Authority
NIH	National Institute of Health
QUT	Queensland University of Technology
SVG	Scalable Vector Graphics
US	United States

1. Introduction

The use of disease and health maps are increasing in popularity. They are powerful tools for communicating sophisticated statistical outputs to non-expert audiences. The design of these communication products is a critical process that must be well informed if the intended message is to reach the intended audience and generated the intended outcome. In order to ensure that these communication products are effective, the visualisation platforms, report measures, and intended audiences must be well researched, and design decisions must align with the communication goals of the overall research project.

The following report provides background information to support the design of the National Cancer Atlas. The sections below provide: an overview of visualization platforms for generating disease maps, both currently available and emerging; a discussion on uncertainty - its importance, sources within the Cancer Atlas and visualisation approaches found in the literature; and finally, an outline of the audience scoping workshops conducted with key stakeholders of the project.

2. Visualisation Platforms for Cancer Atlases

In the Cancer Atlas Grey Literature Review, we outlined the visualisation platforms used to create and publish the cancer atlases identified in the grey literature review. The following section provides more details of these platforms, and considers their application in the context of the National Cancer Atlas. A brief introduction to the javascript libraries of D3.js, leaflet and shiny are also included as examples of emerging popular and versatile tools for creating interactive data driven maps.

In addition to interactive web interfaces, a large number of atlases were published as a pdf (n=9) or a simple infographic (n=1). These atlases had no interactive capabilities, however some (n=2) did provide the option to choose the cancer of interest from a drop down menu, before downloading the image.

Atlases that enabled some level of interactivity were published using the platforms InstantAtlas (n=9), GeoVISTA (n=1) and ESRI (n=2). Custom built maps were built using a range of JavaScript libraries, mapping services and HTML5/CSS frameworks, see Table 2.1 below for more details.

Table 2.1: Visualisation platforms and approaches

Technology Platform/ tools	**Number of Atlases using platform**	**Further Details**
Pdf or infographic	10	n/a
Mapping or Visualisation Platform	Total - 14	
InstantAtlas	9	http://www.instantatlas.com/
Googlemaps based	2	n/a
ESRI ArcMap (in the ArcGIS Desktop suite)	2	http://www.esri.com/software/arcgis
GeoVISTA	1	http://www.geovista.psu.edu/
Custom built	Total - 9	
Custom built - D3.js	2	D3.js - https://d3js.org/
Custom built - D3.js + leaflet	1	Leaflet - http://leafletjs.com/
Custom built – Other	6	A range of JavaScript libraries, mapping services and HTML5/CSS frameworks

2.1 InstantAltas

InstantAtlas was a very popular visualisation platform amongst the maps identified (n=9). This platform provides an easy to use disease mapping tool that takes data in the form of a csv file (observed cancer cases + population data + regional age structure data) and the necessary shape files, and translates this into a disease map. The platform outputs a dashboard that can be published on any website. The dashboard typically contains a disease map, data table, bar chart of cancer rate by region, links to additional resources and is easily customisable to suit the needs of the publisher. All data products within the dashboard are interconnected, and highlighting a region on the map will highlight the corresponding data on the additional graphs and data table in the dashboard. The layout of the dashboard and the map design can be customised to a limited degree. InstantAtlas is a fee based service with a yearly subscription fee of US$1495/year (as at the time of writing, June 2016).

The advantages of the InstantAtlas platform is the limited statistical modelling, html, JavaScript or graphic design skills required to output an interactive web product that shows spatial variation.

An extract from the InstantAtlas website describes the ease of the platform:

"Take data from a spreadsheet, publish an atlas and embed it on your web site in just a few minutes. You don't need any special technical knowledge.
If you can use a Microsoft Excel spreadsheet and an Internet browser then you can publish and maintain InstantAtlas reports."

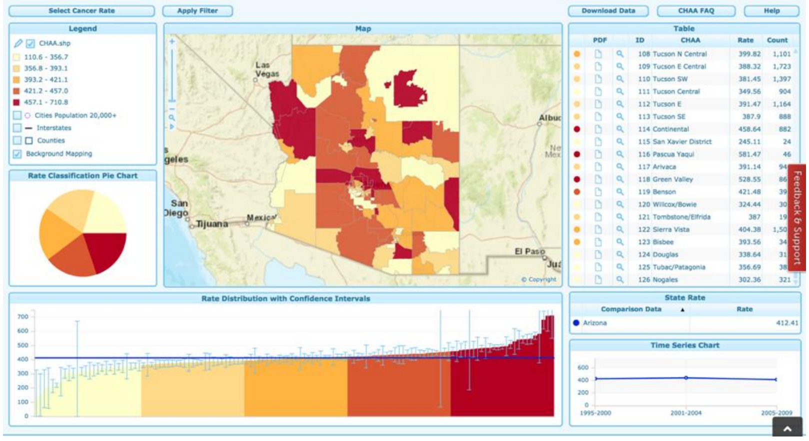
The ease of the platform however does come at a cost, and the final output can be limited in terms of design (can feel a little dated), layout, interactivity and sophistication of statistical methods/reported measure (can only show age adjusted incidence rates). The customisation options and interactive capabilities are limited, and the layout and graphic design are readily identifiable as an InstantAtlas product. Of particular concern are the legend labels, which are regularly very poor, non-intuitive or missing. In addition, the interface is slow to load both initially and also when customizing the view.

InstantAtlas provides a very cost effective solution to developing a web integrated and interactive disease map, particularly in comparison to the costs of hiring consultants or in-house expertise to build a custom product. However, there are many emerging tools that can provide superior design and interactive capabilities, such as D3.js, leaflet, Shiny and ESRI's Story Maps. These new tools and frameworks may not be as simple to use as InstantAtlas, however, improvements are reducing the input required to build a customized, interactive and well-designed data driven disease map.

Instant Atlas Examples

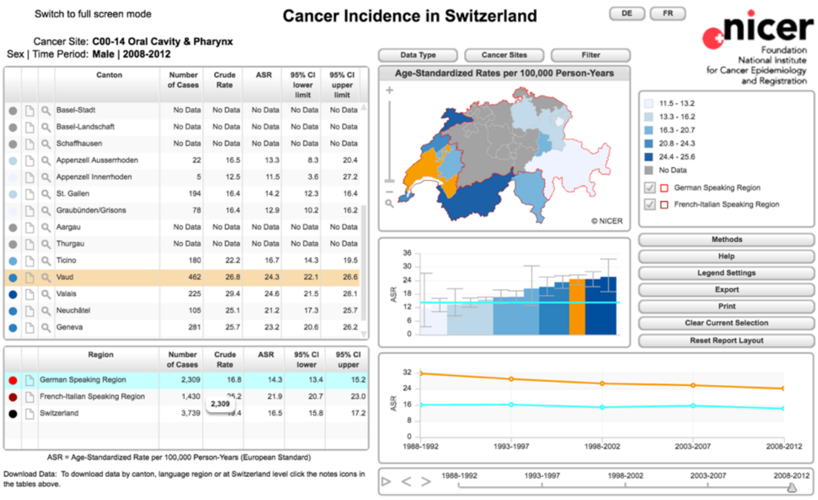
The following figures show four different examples of the InstantAtlas platform, demonstrating the layout and features that are typical of the platform. Additional examples of the platform can be seen in the Cancer Atlas Grey Literature Review, which accompanies this report.

Figure 2.1: Arizona Cancer Rates by Community Health Analysis Areas



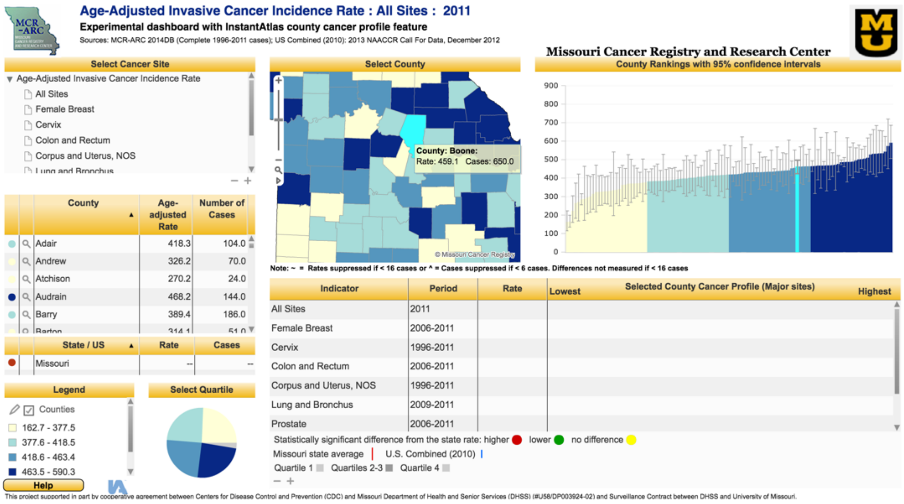
URL: http://www.azdhs.gov/preparedness/public-health-statistics/cancer-registry/chaa/index.php

Figure 2.2: Cancer Incidence in Switzerland



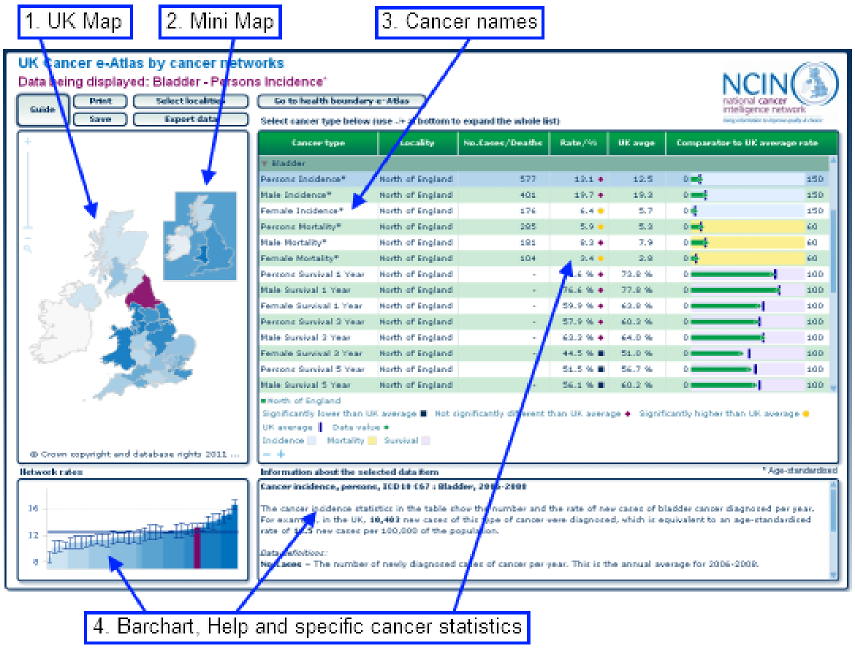
URL: http://www.nicer.org/NicerReportFiles2015-2/EN/report/atlas.html?&geog=0

Figure 2.3: Missouri Cancer Registry: Age Adjusted Invasive Cancer Incidence Rates 2011



URL:http://mcriaweb.col.missouri.edu/IAS/dataviews/report?reportId=13&viewId=3&geoReportId=62&geoId=1&geoSubsetId=

Figure 2.4: Public Health England: NCIN Cancer e-Atlas



URL: http://www.ncin.org.uk/cancer_information_tools/eatlas/

2.2 Custom Built

The cancer atlases that were custom built (n=9) used a range of JavaScript libraries, mapping services and HTML5/CSS frameworks. Table 2.2 below provides an overview of the languages, libraries and frameworks utilised to build the maps. This section seeks to provide information on the commonly used custom approaches, but is not designed to becomprehensive or conclusive.

Table 2.2: Custom built cancer atlases

Map Name	Digital base	# in database & URL
Breast Cancer Mortality in Canada	Jpeg + javascript	#2 http://www.ehatlas.ca/light-pollution/maps/breast-cancer-mortality
Spatio-Temporal Atlas of Mortality in Comunitat Valencia	Unknown	#6 http://www.geeitema.org/AtlasET/index.jsp?idioma=I
Global Cancer Map	Modestmaps + JavaScript + mapbox	#5 http://globalcancermap.com/
CINA+ Online Caner in North America	JavaScript (built from scratch!)	#20 http://www.cancer-rates.info/naaccr/
The Cancer Atlas	JavaScript + Google Maps api	#4 http://canceratlas.cancer.org/data/#?view=map&metric=INCID_ALL_M
Longer Lives	JavaScript + Google Maps api	#11 http://healthierlives.phe.org.uk/topic/mortality
MapNH Health	D3.js + JavaScript + GIS capabilities	#8 http://www.mapnhhealth.org/
Globocan 2012:Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012	D3.js + JavaScript	#3 http://globocan.iarc.fr/Pages/Map.aspx
The Environment and Health Atlas of England and Wales	D3.js + leaflet + JavaScript	#21 http://www.envhealthatlas.co.uk/eha/Breast/

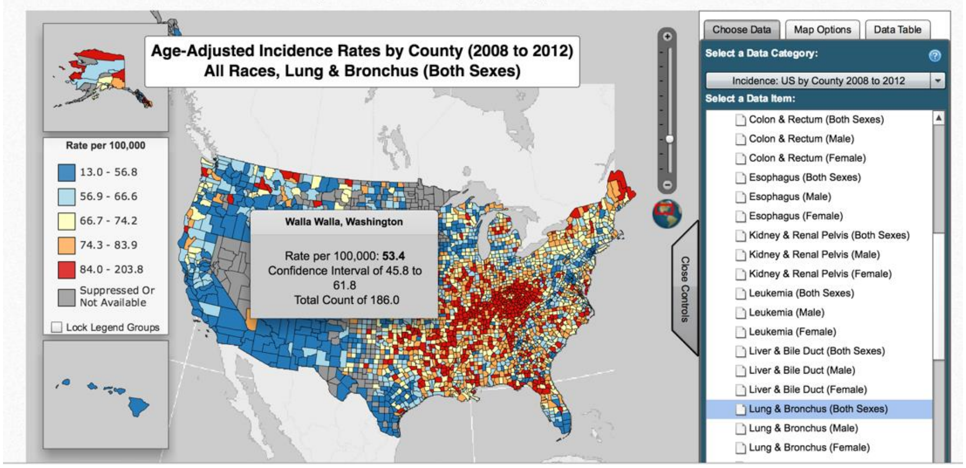
2.3 ArcMap

Within the Cancer Atlas Grey Literature Review (report 1) , two maps were identified that utilised the visualistion tool ArcMap, the *All Ireland Cancer Atlas 1995 - 2007* and the *NCI Geoviewer | NIH GIS Resources for Cancer Research*. While the former utilised ArcMap to create the data visualisation which was published as a pdf, the latter, generated an interactive dashboard where viewers could select the population and disease of interest, as well as other variables (see Figure 2.5 below).

The NCI Geoviewer platform seems a little clunky and dated when interacting with it, and the number of demographic options to choose from feels overwhelming. However, it is worth noting that this map was designed as a research support tool rather than a platform for a general audience. With this in mind the ‘map options’ tab is a useful addition, and could be considered in the design of the National Cancer Atlas. This tab enables the user to customize a limited number of the map design features prior to downloading a pdf version. This could be a very valuable addition for a third party who would like to include the map in their research or communication material. This feature could also be extended to enable other output formats other than pdf, thus enhancing the usability for third party users.

ArcMap has now been superseded by ESRI's Smart Mapping tools within the ArcGIS Online platform ESRI's Story Maps. These two platforms are discussed further in Section 2.5.

Figure 2.5: NCI Geoviewer | NIH GIS Resources for Cancer Research



URL: https://gis.cancer.gov/geoviewer/app/

2.4 GeoVISTA: GeoViz Toolkit

The GeoViz Toolkit was developed from work conducted at the GeoVista Center at The Pennsylvania State University. These tools were developed for the visualization of multidimensional, geographic data. The GeoViz Toolkit is an application version of GeoVista Studio in which a sample of these components were adapted to a more “user-friendly” environment. This toolkit was developed by Frank Hardisty, Aaron Myers, and Ke Liao at the University of South Carolina.

Users can select a range of data products to include in the final disease map dashboard, and all of the components are linked. For example, if you select a different set of variables in one component then that change will be reflected in all other components. The final dashboard is rendered in flash and the interactive features are very fast.

Unfortunately, the platform appears to only cater for US geographies, so has limited application for the National Cancer Atlas unless further development work is applied.

2.5 Emerging Platforms

2.5.1 ESRI – Story Maps

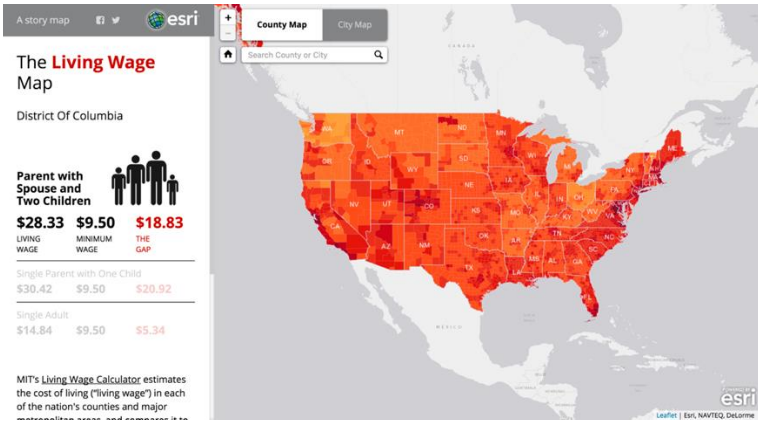
Although none of the identified cancer maps used the ESRI's Story Maps platform, the *NIH GIS Resources of Cancer Research* (seen in Figure 2.5) was generated using its ArcGIS predecessor ArcMap, and it is emerging as a popular tool for creating interactive maps. As part of the ArcGIS Online offering, ESRI Story Maps combines authoritative maps with narrative text, images, and multimedia content. Story Maps provides a platform for building a data driven story around spatial variation that integrates data and content from a range of formats including video, images, text, interactive web graphics and maps. An ArcGIS Online account is required to use this platform.

This platform can interface with tools such as leaflet (see Section 2.5.2) to enhance the interactive capabilities of the map, and has a wide range of templates that can be used to build a custom story map. Examples are shown in Figure 2.6 and 2.7.

Further info: http://storymaps.arcgis.com/en/how-to/

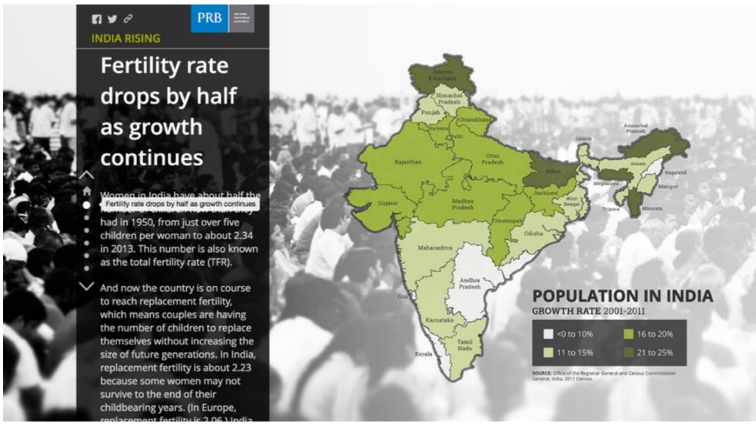
Example: https://storymaps.arcgis.com/en/gallery/#s=0

Figure 2.6: The Living Wage Map



URL: http://storymaps.esri.com/stories/2015/living-wage-map/

Figure 2.7: India Rising



URL: http://www.arcgis.com/apps/MapJournal/index.html?appid=d9a6a166ac6e409a9a1ae6cbb4709056

2.5.2 Leaflet

Leaflet (http://leafletjs.com/) is an open source JavaScript library for building mobile-friendly interactive maps. It's used by websites ranging from The New York Times and The Washington Post to GitHub and Flickr, as well as GIS specialists like OpenStreetMap, Mapbox, and CartoDB. Leaflet can be used with both the statistical software language R and D3.js to build interactive web based maps.

Further info: http://leafletjs.com/

Examples:

1. New York Times - http://www.nytimes.com/projects/elections/2013/nyc-primary/mayor/map.html
2. The Washington Post - http://www.washingtonpost.com/sf/local/2013/11/09/washington-a-world-apart/
3. GitHub - https://github.com/blog/1528-there-s-a-map-for-that
4. ESRI Story Maps + Leaflet - http://storymaps.esri.com/stories/2015/living-wage-map/

2.5.3 Within the statistical software language R

Shiny + Leaflet

Shiny is a web application framework for R. Shiny enables interactive web applications to be built from within the R environment. This powerful tool enables statisticians and data scientists to create interactive web applications without needing expert skills in html or JavaScript. When combined with leaflet for R, Shiny enables statisticians and data scientists to create interactive data driven maps.

An advantage of using Shiny, compared to a platform that uses pre-processed data (such as D3.js), is whenever the map is displayed the data is processed by the server running R, therefore the web interface responds to changes in the underlying data.

Further info: https://rstudio.github.io/leaflet/

Examples: http://shiny.rstudio.com/gallery/superzip-example.html

Limitations: Hosting - The R server needs to be running somewhere to drive the Shiny interface, this is not something that is normally installed on a web server.

2.5.4 D3.js + Leaflet

D3.js is a JavaScript library for manipulating documents based on data, and helps you bring data to life using HTML, SVG, and CSS. D3's emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, and combines powerful visualization components with a data-driven approach to the Document Object Model (DOM) manipulation.

More information: https://d3js.org/

Examples:

1. D3.js + leaflet - https://www.infino.me/mortality/usmap
2. Other - https://github.com/d3/d3/wiki/Gallery

2.5.5 Epiphanee

Epiphanee is a powerful spatial query and visual analytics tool which uses sophisticated privacy filters to maintain anonymity compliance while enabling users to query the data. The platform is capable of visualising spatio-temporal data, which can be embedded within a website. The platform enables users to submit a range of queries without providing access to private or sensitive aspects of the data. The data is dynamically linked and automatically updates the rendered map.

More Information: http://www.crcsi.com.au/impact/visualisation-and-analysis/

2.6 Additional platforms

There are a range of additional data visualisation platforms that are worth noting. While these platforms are not specifically targeted towards the visualistion of spatial data they may be valuable data visualistion tools to consider.

1. Qlikmaps - http://www.qlikmaps.com/
2. Tableau - http://www.tableau.com/
3. Microsoft Power BI - https://powerbi.microsoft.com/en-us/

3. Communicating Statistical Uncertainty for Health Maps

Health maps can be very powerful tools for communicating the output of sophisticated spatial statistical analyses to non-expert audiences and decision makers. However, rendering these statistical insights into a visual map can make them appear more certain than they really are. The results displayed in health and disease maps are usually the best estimates that are available at the time, but they are still estimates, and information about the reliability, accuracy, or precision of these estimates is rarely included.

Communicating statistical uncertainty to the users of these maps is a complicated design and communication challenge with limited, if any, accepted standards. Uncertainty does not feature strongly in the cancer maps identified in the Cancer Atlas Grey Literature Review, nor are tools for including uncertainty readily available in the platforms described in Section 2. However, if decision and/or policy makers are to make the best decisions possible on the available information, finding ways to communicate scientific uncertainty will be essential.

3.1 What is Uncertainty

Uncertainty is a daily fact of life, it is present in all areas of life and science, and generally is a measure of belief about a statement. In the scientific literature, uncertainty is not a simple, clear, or well-defined concept, and there are many different interpretations found across knowledge domains (MacEachren, 2005). Within the scientific literature, uncertainty can be used to refer to: different views, imprecision, error, subjectivity, non-specificity, a lack of knowledge, or a state of being (Aerts, Clarke, and Keuper, 2003; Pang, Wittenbrink and Lodha, 1997; Deitrick & Edsall, 2008; Thomson *et al.*, 2005; Han *et al.*, 2011). Various attempts have previously been made to harmonize the disparate literature on uncertainty (Morgan & Henrion ,1990; Regan *et al.*, 2002; Walker *et al*., 2003; Kujala *et al.*, 2012), but there is still no single agreed upon use or meaning of the term.

Kujala *et al*., (2012) classifies uncertainty into 3 main categories: (1) linguistic uncertainty, where one term can be interpreted in several ways, (2) Human decision or behavioral uncertainty, which is defined as the uncertainty about worldviews, objectives and stakeholders, and (3) epistemic uncertainty, which is uncertainty about facts. We propose two extensions of these broad categories, in order to more accurately cover the different definitions and types of uncertainty.

Firstly, we propose that a fourth category, ignorance (or the experience of being uncertain), as defined by Kahneman & Tversky (1982), should be added to this list. Ignorance is defined as the internal uncertainty that is experienced when a person is unsure, and is present in the statements "I hope I spelt her name correctly", " I think Mt. Blanc is the highest mountain in Europe" or "I'm not sure which way to go". The experience of being uncertain implies a consciousness or awareness of one's lack of knowledge, and in this sense uncertainty is a form of metacognition, a knowing about knowing (Flavell, 1976) or a subjective perception of ignorance.

Secondly, we find the definition of Kujala et al.'s (2012) third category, epistemic uncertainty, is insufficient for the purpose of this research. Instead, we replace epistemic uncertainty with scientific uncertainty as defined by Han *et al.* (2011), which includes both aleatory and epistemic uncertainties. Aleatory uncertainties are a result of the underlying randomness within the model, or processes that are being modeled. They cannot be reduced by gathering more information or measurements and are a result of the fundamental irreducible randomness or indeterminancy of natural events. While epistemic uncertainties are presumed to be due to a lack of knowledge, and can be reduced by gathering more data or refining the model.

Our final classification of uncertainty is thus:

1. Linguistic imprecision
2. Human decision/behavioural uncertainty
3. Scientific uncertainty
    - Aleatory
    - Epistemic
4. Ignorance/experience of being uncertain or not knowing

In this study, we focus on this redefined third category of uncertainty, scientific uncertainty. Unless explicitly stated otherwise, we use the term uncertainty from here on to refer to scientific uncertainty.

We make a special note, at this point, about the first category of uncertainty listed above, linguistic uncertainty or imprecision. While scientific uncertainty is the focus of this research, we are investigating the communication of scientific uncertainty, therefore, it is important to also be aware of linguistic uncertainty as it is an important consideration in communication design. Linguistic uncertainty is not a source of scientific uncertainty, nor is it a focus of the uncertainty visualization sections of this report, but it is important to consider terminology and labels and how these may be interpreted differently by different audiences.

We also note that in this research we do not consider uncertainty in the context of quantum mechanics, as defined by the Heisenberg's uncertainty principle (Hughes, 1989), but focus on scientific uncertainty as related to empirical quantities, where empirical quantities are properties of the real-word that can, in principle be measured to some level of accuracy, (e.g. birth weight, height, rainfall) (Begg, 2014).

3.1.1 Risk vs Uncertainty

The terms "uncertainty" and "risk" can often be confused or erroneously used interchangeably. Within a modelling context, uncertainty refers to any deficiency in the modelling process, methods or data that is not definite, not known or not reliable. This means that some relevant information about the estimates, or the model outputs, is not known or un-knowable (Thunnissen, 2003).

Similar to uncertainty, risk has more than one definition within the academic literature. For example, Bedford & Cooke (2001) define risk as the possible impact or outcome of an uncertain situation or problem. While Knight (1921), describes risk as the calculable and therefore controllable part of all that is unknowable, and the remainder is uncertainty - incalculable and uncontrollable.

We find the later example does not align as well with our definition of uncertainty as the former, however, regardless of which definition of risk the reader agrees with, it is important to note that we are not discussing risk within this work. We acknowledge that uncertainty can be an important input in evaluating risk, but we emphasize that they are different concepts, and leave risk to a separate discussion.

3.1.2 Variability vs Uncertainty

Uncertainty and variability are also often confused in practice, and while variability is an important source of scientific uncertainty, the terms are not interchangeable. Variability can be thought of as a feature of the observed population, while uncertainty exists in the context of an estimate of the true value of a single event or quantity (Begg, 2014).

Variability attempts to describe a characteristic or feature of a population, and is a descriptive feature of a data set of observed values. Variability arises when multiple measurements of an event or phenomenon are observed, and is a natural feature of the population being studied. Take for example, the birth weight of babies born in Australia, in 2015. The weight of each baby is different, and the range from the smallest to the largest birth weight is due to natural variation. This variation gives rise to a natural and expected birth weight range for babies from a specific population. Equipment accuracy and measuring methodologies can influence the variability of observed data, but variability cannot be removed by improving methodologies and measurement accuracy or collecting more observations. Variability is considered predominantly due to irreducible natural variation.

In contrast, uncertainty can be considered as the inaccuracies or indeterminacies associated with predicting the weight of the next baby born in a population. Variability contributes to our ability to accurately predict the birthweight of the next baby born. Variability is a source of uncertainty, and is sometimes used as a measure of uncertainty, but the terms cannot be used interchangeably.

Much of the confusion between uncertainty and variability arises due to the fact that distributions are used to describe both. Uncertainties are often quantified using probabilities and probability distributions, which are assigned using information or evidence we have about what the true value or most likely future value, might be. Variability is quantified using a frequency distribution derived from measurements or observations, ie. data. The confusion is often compounded by the fact that frequency distributions, derived from an observed population, are often used to inform a probability distribution in order to predict a future event or make assumptions about a similar population.

Variability statistics can be very valuable for informing or assessing probabilities, with the usefulness depending on how similar the processes that produced the observations are to the processes that produced (or will produce) the event which we interested in.

Variability = Observed, frequency distribution, source of uncertainty.
Uncertainty = Predict unknown, probability distribution, informed by frequency distribution.

3.2 Why is Uncertainty Information Important for Decision-Makers?

"All models are wrong, but some are useful"
(George E. P. Box)

Planning for the future, allocating resources and evaluating multiple options are all decision making processes that require the decision-maker to do the best they can with the knowledge they have available at the time. Acknowledging uncertainty can clarify the reliability, accuracy and precision of available information. It can be a powerful tool that supports reasoning and enables more informed decision making while using all available information. In the emerging big data realm with increasingly complex problems, the valuable information from uncertainty can be dangerous to ignore, and will provide a competitive edge to those that can utilise it appropriately.

This famous quote above, from the statistician George E. P. Box, succinctly expresses that we cannot know something completely, but incomplete knowledge or information is still valuable. If we accept that we can never know everything completely, then everything we do know we know only partly. Uncertainty helps clarify how much we know, it is an indicator of the reliability and accuracy of our knowledge, and by acknowledging, quantifying and communicating uncertainty we transform a lack of knowledge into an informative piece of information that can be acted upon.

"To be uncertain is uncomfortable but to be certain is ridiculous"
Chinese Proverb.

In most decision making contexts it is not possible to wait until perfect information is available before making a decision, often action is required 'now'. The information obtained from uncertainty can support better decision-making in three main ways. Firstly, it can ensure that incomplete information is applied appropriately. Dangers arise when information is presented, or appears, more certain than it is. This is particularly important in disease mapping, because the act of rendering a statistical estimate on a map can make the information appear more certain than it really is. Secondly, uncertainty can guide future research and inform where new or updated information is most needed. Thirdly, in contexts that have a lot of unknowns, quantifying what is known and unknown can relieve the internal anxiety and choice paralysis that can arise.

As scientists attempt to model and research increasingly complex systems and relationships, and the availability of data increases, statistical uncertainty becomes more than a dismissible piece of metadata. Uncertainty becomes a valuable source of information that can support reasoning and enable more informed decision-making. Ignoring and/or not understanding uncertainty information can result in misinterpretation of model outputs, substandard decisions that do not leverage all available information, or information being disregarded completely due to too much uncertainty.

3.3 Why Now?

In the past uncertainty has been ignored, neglected and suppressed because its presence has led to information being discarded, users being confused or data without uncertainty being preferred because it appeared to be of a higher quality, when in fact it may not have been.

Uncertainty was also in part ignored because of a lack of tools, or understanding around how to handle it, how to use it in decision making, and how to communicate it to the decision-maker or non-expert (Lindley, 2006).

Despite these challenges, there is a change taking place in attitudes towards uncertainty. Although this change may not have fully reached the general public or the non-expert decision maker, it is on its way. Uncertainty used to be considered information only to be shared with and between scientists, but is now being recognised as a critical element of scientific output that must be communicated to the consumers of scientific insights.

This change is driven by a range of factors, including:

- Increase in the availability of data and the emergence of the 'big data revolution'.
    - Growth in the awareness of big data in popular science and traditionally, non-data driven industries, has increased the number of non-experts using data driven insights to make decisions. Increased markets demanding access to the 'big data revolution'.
    - Technological innovations are reducing the cost of collecting data, therefore increasing the volume of data available. But also, the emergence of lower quality 'cheap' sensors is increasing the volume of 'dirty data'.
    - Increase in 'dirty data' which contains higher rates of error and missing data. The emergence of affordable lower quality data collection technologies and the increase of non-experts collecting data gives rise to the collection of data which contains higher rates of errors and missing data.
- Users of data driven insights outnumber the producers (Enserink, 2013). Therefore the uncertainty information much be communicated explicitly to the user, rather than being understood implicitly by the producer.
- Scientific investigation of increasingly complex systems, processes and relationships which may contain irreducible uncertainty. This is also contributing to a greater tolerance of uncertainty in scientific output: as the complexity of the system under investigation increases, the ability to output a deterministic result decreases.
- Technological advances in computing power
    - Greater uptake of Bayesian and simulation based methods that explicitly consider uncertainty.
- Philosophical change in scientific attitudes in many traditional fields of systems and decisions science are moving from deterministic modelling to methods that explicitly take uncertainties into account (Grubler *et al*., 2015).
- Development of better tools to handle, represent and communicate uncertainty.

The need to address uncertainty is increasingly recognised within the academic literature across a wide range of fields. The following list provides some examples from the literature

1. **Natural resource management** - understanding that average processes is often not sufficient to make a robust decision, and decision-makers are increasingly interested in understanding the uncertainties of the models (O'Hagan, 2012; Uusitalo et al, 2015).

2. **Conservation and environmental science** - recent publications argue for explicit assessments of uncertainty in environmental data and models as a necessary, although not sufficient, condition for balancing uncertain scientific arguments against uncertain social, ethical, moral and legal arguments, in managing environmental systems (Brown, 2004 and Uusitalo *et al*., 2015).

3. **Health care** – uncertainty, including scientific uncertainty is a matter of preeminent concern given the growing emphasis on evidence-based medicine and on informed and shared decision making (Han *et al.*, 2011). Also, Politi *et al.*, (2007) provides an insightful window into how the health industry is trying to grapple with the complex relationship between evidence based practice, informed consent and uncertainty. In practice, many clinicians struggle with communicating and understanding how uncertainty should impact the informed consent process.

4. **Rare and extreme events -** Harrower & Street (2003) state that *"If there is one thing that defines and limits our efforts to better understand extreme and rare events it is uncertainty."*

5. **Spatial Analysis -** *"While there has already been considerable research undertaken to develop models of spatial data error and uncertainty, there is an additional requirement for the results of these models to be effectively conveyed to users."* (Hunter and Goodchild, 1996).

6. **Policy -** the past decade has seen a growing recognition that policies that ignore uncertainty about technology and the physical world, often lead in the long run to unsatisfactory technical, social, and political outcomes. Recent growth in interest, understanding, and technical skill in the field of risk analysis and assessment has worked to promote this change. The fact that risk inherently involves chance or probability leads directly to a need to describe and deal with uncertainty (Morgan & Henrion, 1990).

7. **Systems and Decisions Science -** Grubler *et al.* (2015) argues for the need for a *'paradigm shift in many traditional fields of systems and decision science: moving from deterministic modelling (with hopefully extended sensitivity analyses) and its (futile) quest for "optimality" to the concept of "robust" decision making that takes uncertainties explicitly into account, transforming optimality conditions from singular decision variables/criteria to whole ensembles, or portfolios of response options, that in their combination constitute optimal hedging strategies with regards to uncertainties'.*

We conclude this section with an extract from Deitrick & Wentz' 2015 publication titled *Developing Implicit Uncertainty Visualization Methods Motivated by Theories in Decision*

Science. This extract succinctly summarises current difference views towards uncertainty between researchers and policy-makers.

> *"Public policy decision makers, defined here as individuals with useful decision-making knowledge or the ability to enact a policy, understand that uncertainty is an inescapable component of decision making. Similarly, geographic information systems (GIS) and geovisualization researchers (referred to as researchers here) recognize the importance of identifying and evaluating uncertainty in analysis and outputs for decision support. Nevertheless, specific visualization methods and tools for incorporating uncertainty into GIS are not widely used or requested by decision makers. Moreover, research indicates that decision makers often view GIS uncertainty visualizations and tools as a constraint to making decisions, which might lead them to avoid solutions that employ uncertain information or to overly rely on the results of prior similar decision tasks. Because there is agreement between decision makers and researchers that uncertainty is important, yet disagreement in how to incorporate it into decision support, we see this as a discrepancy between the way decision makers and researchers conceptualize uncertainty in decision problems."*

3.4 Sources of Uncertainty

I could pretend that the answer is simple, and it would make you happy, but I wouldn't be being very honest.

(Royal Statistical Society, President's address, 2015. Peter Diggle)

The presence of uncertainty is well recognized, and its importance is increasingly appreciated by policymakers and decision-makers. However, a clear agreed upon definition of the different sources and characteristics of uncertainty is difficult to find (Walker *et al*., 2003). In this section we outline an appropriate framework for the context of the National Cancer Atlas, and provide a list of potential alternative frameworks.

There have been many attempts at defining uncertainty typologies in the literature and these are both varied and overlapping. These include, but are not limited to Regan *et al.*, (2002 & 2008), Kahneman & Tversky (1982), Morgan & Henrion (1990) Walker *et al*., 2003, Skinner *et al*., 2014 and Han *et al*., (2011). The differences between these taxonomies are predominantly due to context (or domain) and scope (or detail). Some are very generic (Kahneman & Tversky (1982); Morgan & Henrion, 1990; Walker et al., 2003; Funtowicz & Ravetz (1990)), while others are domain specific (ecology & conservation biology - Regan et al., 2002 & 2008; environmental risk assessments - Skinner *et al*., 2014; hydrology - Beven *et al.,* 2015; health - Han et al., 2011; intelligence analysts & geospatial data - Thomson *et al.*, 2005; medical imaging - Ristovski *et al*., 2014; software engineering - Ramirez *et al*., 2012).

For the purposes of the National Cancer Atlas, and for disease mapping more generally, we propose the taxonomy of uncertainty sources as detailed in Table 3.1.

Table 3.1: Taxonomy of uncertainty for disease mapping

Location	Source	Aleatoric vs Epistemic
The Data (inputs)		
	Measurement error (random error or statistical variation)	Epistemic
	Systematic error (subjective judgement or bias)	Epistemic
The Model (Knowledge)		
	Model structure	Epistemic
	Model parameters	Epistemic
	Approximation	Epistemic
	Disagreement	Epistemic
	Inherent randomness	Aleatory
Other		
	Natural variability	Epistemic
	Linguistic uncertainty	n/a

This taxonomy is a combination of the detailed taxonomy defined by Morgan & Henrion (1990), including 'model uncertainties' as defined by Regen *et al.*, (2002), with the broad framework defined by Funtowicz and Ravetz (1990). Funtowic and Ravetz (1990) categorises uncertainties as either:

- Data uncertainties - arising from the quality or appropriateness of the data used as inputs;
- Modelling uncertainties - arising from an incomplete understanding of the modelled phenomena or from numerical approximations used in mathematical representations of processes; or
- Completeness uncertainties - referring to all omissions due to a lack of knowledge.

We have found it more useful to use 'Other' instead of 'Completeness Uncertainties'.

The terms in Table 3.1 are defined as follows.

1. ***Measurement error (random error or statistical variation)*** is the most studied and best understood kind of uncertainty and arises from random error in direct measurements of a quantity. Imperfections in the measuring instruments and observational techniques will inevitably give rise to variations from one observation to the next. The resulting uncertainty depends on the size of the variations between observations and the number of observations taken. The measurement error can be estimated by statistical methods, and there are a variety of well-known techniques for quantifying this uncertainty, such as standard deviation and confidence intervals.

2. ***Systematic error* (*subjective judgment or bias)*** in measurements or observations results from a bias in the sampling or measuring equipment, and is more difficult to quantify, or even notice. If systematic error goes unnoticed, it may have cumulative effects in the models that are built into the data. Systematic error can also be considered to be subjective judgment or bias as the only way to deal with systematic error is to recognize a bias in the experimental procedure and remove it. How the bias is removed is purely up to subjective or scientific judgment. Very little can be done to quantify or address systematic error, due to its unknown nature. If we knew there was systematic error, we would address it, but we are often not aware it is there: these are the unknown unknowns.

3. ***Variability (natural variation)*** occurs naturally in a measured quantity over time and space. For example, the weight of normal term babies born in Queensland in 2015 will vary. Variability is not uncertainty in itself, but a characteristic of a population of observations. However, variability is a source of uncertainty when we use observed information from a population to predict or estimate an unobserved event or measured quantity. For example, the natural variability in Queensland birth weights in 2015 gives rise to uncertainty when this information is used to prediction of the birth weight of the next baby born in Queensland.

   Variability is a source of aleatory uncertainty and cannot be reduced by collecting more observations. However, understanding the factors that underlie variation in an observation, can expand our understanding of the system sufficiently enough to reduce the uncertainty that arises due to natural variation. Refer to Section 3.1 for further discussion on variability vs uncertainty.

4. ***Model uncertainty*** arises because models are always abstractions of the natural system. Some less important variables and interactions are left out, and the shapes of the functions are always abstractions of the real processes. The current understanding of the process may be incomplete, and the shapes of the functions and their parameter values are only best estimates. Uncertainty of the *model parameters* can be accounted for in probabilistic models, with careful consideration of the range of possible values and their probabilities. While uncertainty about the *model's structure*, i.e. uncertainty about cause-and-effect relationships, is often very difficult to quantify.

5. ***Approximation*** gives rise to uncertainty by introducing simplified abstractions of the real-world system into the model. Spatial and temporal resolutions of a model are approximations, and so is the resolution in terms of time intervals and grid size. Approximations are used due to a lack of knowledge about a specific feature of one aspect of the system being modelled or due to computational limitations.

6. ***Disagreements*** give rise to uncertainty when researchers must select between statistical, computational and mathematical methods and techniques that may not be agreed upon within the scientific community. Disagreement may also be a source of uncertainty due to differing interpretations of scientific evidence - knowing differently among decision makers.

7. ***Inherent randomness*** can be thought of as being innate. However well we know the process and the initial (starting) conditions, we cannot be certain of what the outcome will be. Randomness can often be quantified very well, and is easy to deal with in probabilistic models. There is still much debate in the literature about whether inherent randomness is in principle reducible or not (epistemic vs aleatoric uncertainty), a phenomenon of the natural world or a by-product of our lack of knowledge about a system or relationship, and its starting conditions.

   Within a quantum mechanics framework, inherent randomness arises from the Heisenberg indeterminacy and cannot be reduced. However, others still subscribe to Einstein's dictum, that "*God does not play dice*", and thus, all apparent randomness is actually due to a lack of knowledge of either the system or its initial starting conditions.

In this report we take a personalist view of randomness - a value is considered random if you do not know of any pattern or model that can account for its variation. In this view, randomness, like probability, is a function of the knowledge available. A quantity may be legitimately random to one person, but deterministic to another who knows its underlying generating process. A main feature of inherent randomness is that it is not practicably reducible.

"We don't understand how it works, so we will assume that nature doesn't either and so it behaves randomly "
E. T. Jaynes. Probability Theory: The Logic of Science. Cambridge University Press. 2003.

8. ***Linguistic Uncertainty (linguistic imprecision)*** arises because both natural and scientific language can be interpreted in several ways, or an event is ill-defined. Linguistic uncertainty can be classified into five distinct types: context dependence, ambiguity, indeterminacy of theoretical terms, and under-specificity. All of these uncertainties arise in natural and scientific language, and can impact the interpretation and application of scientific insights to real world decision making. Of these, vagueness is the most important for practical purposes.

3.4.1 Alternative Frameworks

There are many additional frameworks for classifying sources of uncertainty found in the literature. We have selected above that which is most appropriate for empirical quantities, and it is beyond the scope of this report to detail all alternative frameworks. Table 3.2 provides a non-exhaustive list of these alternative frameworks and their references.

Table 3.2: Uncertainty taxonomies found in the literature

Typology classes	Reference
Internal vs external	Kahnemen & Tversky (1982) Han *et al.*, 2011
Ignorance	Lipshitz & Strauss (1997)
The scientific pipeline (from data acquisition, to transformation, and modeling, and then data visualization)	Pang *et al.*, (1997) Johnson & Sanderson (2003) Brodlie et al., (2012)
Location, level and nature	Walker et al., (2003)
Context, inputs (data) and model	Refsgaard et al., (2007)
Data uncertainties, modelling uncertainties and completeness uncertainties	Funtowicz and Ravetz (1990)

3.4.2 Uncertainty in the National Cancer Atlas

The above defined Taxonomy can be used as a diagnostic tool to analyse the uncertainty within the National Cancer Atlas. Specific sources of uncertainty within the projects' data and methods should be identified, as well as sources of natural variation and linguistic imprecision. For each specific uncertainty source identified, it should be evaluated in terms of reducibility, actions required to address, impact of addressing, impact on decision-maker and recommended actions. This then provides a mechanism for communication between designers and researchers to discuss which sources of uncertainty should be targeted for communication and which should be targeted for reduction or future research. Not all sources of uncertainty are useful for communication or appropriate targets for reduction. Identifying those that cannot be reduced efficiently or practicably, or those that have not influence on a decision making process, guides communication design and future research design.

3.5 Uncertainty Visualisation

Visualization is a proven channel for effectively and efficiently communicating data driven insights to non-expert audiences. The human visual system is a very high-bandwidth channel to the brain, with a significant amount of processing occurring in parallel and at the pre-conscious level. Data visualization enables an audience and a communicator to take advantage of the highly evolved and sophisticated analysis capabilities of the human visual system (Munzner, 2014).

As uncertainty has increased in importance across many academic domains, a growing body of literature, has emerged that specifically explores the visualization of uncertainty information (Mathews *et al.*, 2008; Brodlie *et al.* 2012; Potter *et al.*, 2012; Zuk & Carpendale, 2006; Pang *et al.*, 1997; Johnson & Sanderson, 2003; Johnson, 2004; Evans, 1997; Wittenbrink *et al.*, 1996; Aerts *et al.*, 2003; Sanyal *et al.*, 2009).

3.5.1 Geospatial data

A wide variety of methods have been developed to graphically represent uncertainty in geospatial data (Brodlie *et al*., 2012; Slocum *et al.*, 2003; MacEachren *et* al., 2005; MacEachren, 1992). Pang *et al.*, (1997) suggests these methods can be categorized into three groups: overloading, side-by-side comparison, or seamless integration. Overloading and side-by-side comparisons are also referred to as bivariate maps and map pairs respectively (MacEachren, 1992). These three broad visualization approaches are defined as:

1. **Overloading (Bivariate Maps)** - report data and the associated uncertainty information within one map. Overloading is an approach that augments a base visualization technique with an uncertainty visualization technique but the data and uncertainty information are clearly separable. This is probably the most popular mechanism for uncertainty visualization.

2. **Side-by-side comparison** (**Map Pairs**) - two similar maps presented side by side. One map shows the data, and the other shows the associated uncertainty.

3. **Seamless integration** - the data and the uncertainty are displayed in a unified rendering. Unlike the overloading approach in which uncertainty is superimposed on the graphical representation of the dataset, the seamless integration approach directly includes (i.e., integrates) the uncertainty in the visualization rendering.

Among these three types of methods, research suggests that bivariate maps are the most popular approach. Multiple studies have found this method to be easier to interpret and enhances the user performance compared to separate maps (Kubicek & Sasinka, 2011; Viard *et al.*, 2011; Evans, 1997). These studies did not however compare the type of side-by-side comparisons that we saw in the internet published cancer maps (Cancer Atlas Grey Literature Review 2016), where confidence intervals are included in a supplementary graph of estimate vs region, rather than uncertainty information shown on a version of the main map.

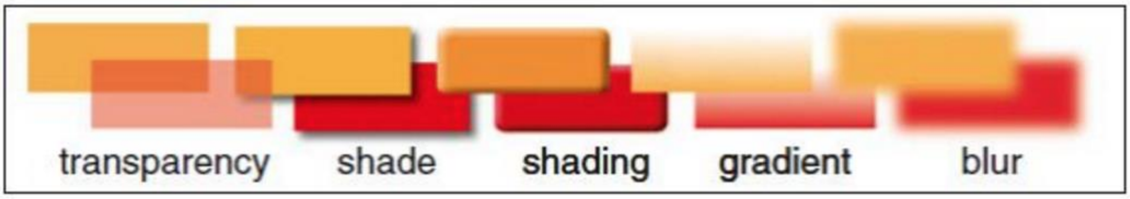
3.5.2 Visually Coding Uncertainty

There are many approaches that have been developed to visually represent uncertainty information. Bertin (1981) suggested location, size, value, texture, color, orientation, and shape. MacEachren's 1992 paper suggested edge crispness (fuzziness), fill clarity, fog, and resolution as valuable approaches while Gershon (1992) suggested boundary (thickness, texture, and color), transparency, animation, and extra dimensionality. The following is a non-exhaustive list of visual variables or symbols found in the literature for representing uncertainty:

- Fuzziness
- Error bars and credible/confidence intervals
- Transparency
- Heat maps
- Colour mapping
- Grid Annotation Lines
- Probability distribution curves
- Boxplots and interquartile range
- Glyphs

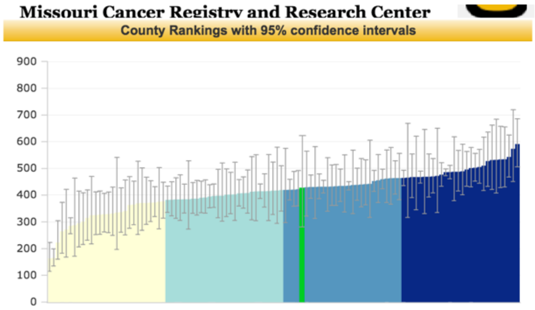
Figures 3.1 to 3.8 show a range of uncertainty visualisation examples.

Figure 3.1: Visualisation methods for categorical data



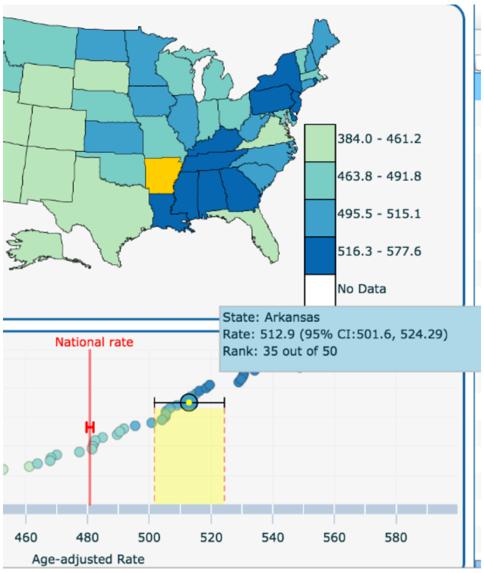
Source: Figure 2 in Vulling *et al.*, 2013.

Figure 3.2: Error bars I



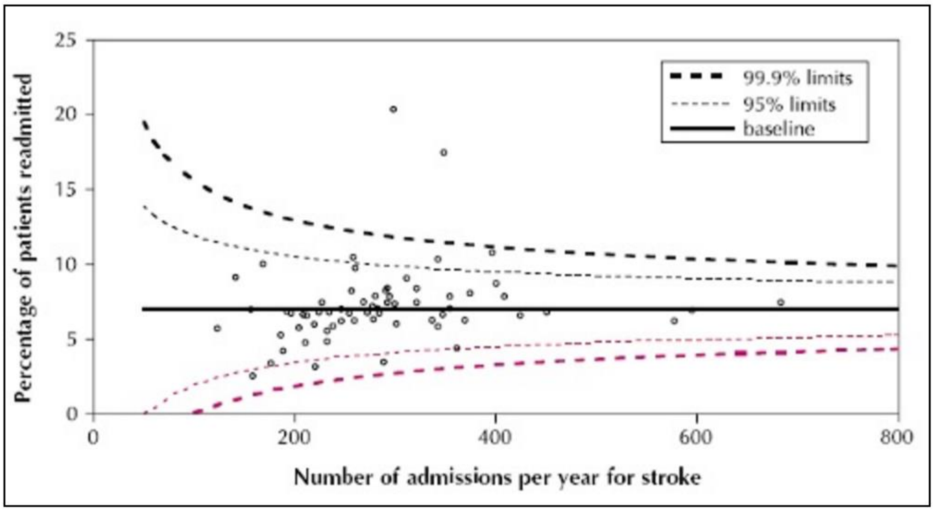
Source: [*http://mcriaweb.col.missouri.edu/IAS/dataviews/report?reportId=13&viewId=3&geoReportId=62&geoId=1&geoSubsetId*](http://mcriaweb.col.missouri.edu/IAS/dataviews/report?reportId=13&viewId=3&geoReportId=62&geoId=1&geoSubsetId)=

Figure 3.3: Error bars II



Source: https://nccd.cdc.gov/DCPC_INCA/

Figure 3.4: Funnel Plots



Source: David Spiegelhalter, Medical Research Council Biostatistics Unit
Shown in ERPHO (2003). Quantifying performance: using performance indicators.
http://www.erpho.org.uk/Download/Public/6990/1/INPHO%204%20Quantifying%20performance.pdf (PDF)

Funnel plots (Figure 3.4) are a powerful way of visualising uncertainty and confidence intervals, particularly where performance data is compared against targets.

Figure 3.1: Grid annotation lines

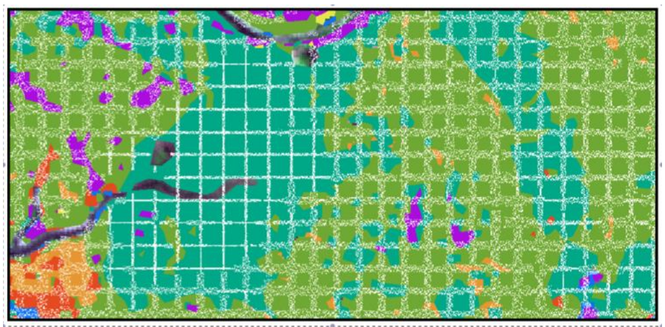


Figure 1. Noise annotation lines representing classification uncertainty on a vegetation land cover map. The local width of the noise grid indicates the degree of uncertainty (larger noise width = higher degree of uncertainty).

URL: http://www.exegetic.biz/blog/tag/scatter-plot/

Grid annotation lines (Figure3.5) were first described as "procedural annotations" by Cedilnik and Rheingans (2001). In this method a grid is placed on the map and grid lines are distorted locally to represent the uncertainty in the data object that lies below the grid. Grid Annotation Lines could be used to represent any type of uncertainty. The grid in Figure 3.2 represents noise. Usability of this method has been tested by Kinkeldey *et al*., (2014).

Figure 3.2: Transparency vs colour mapping

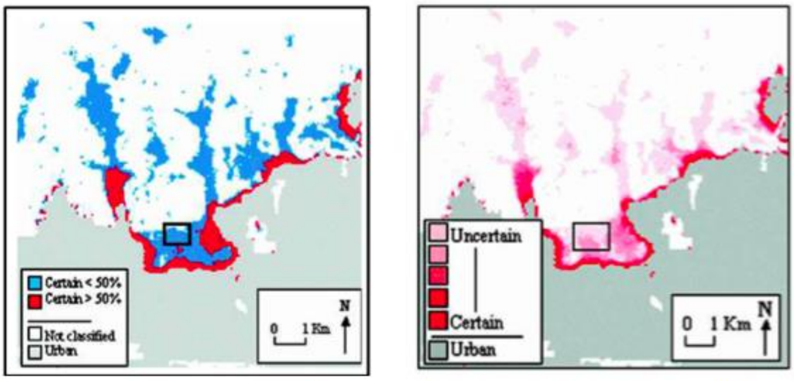
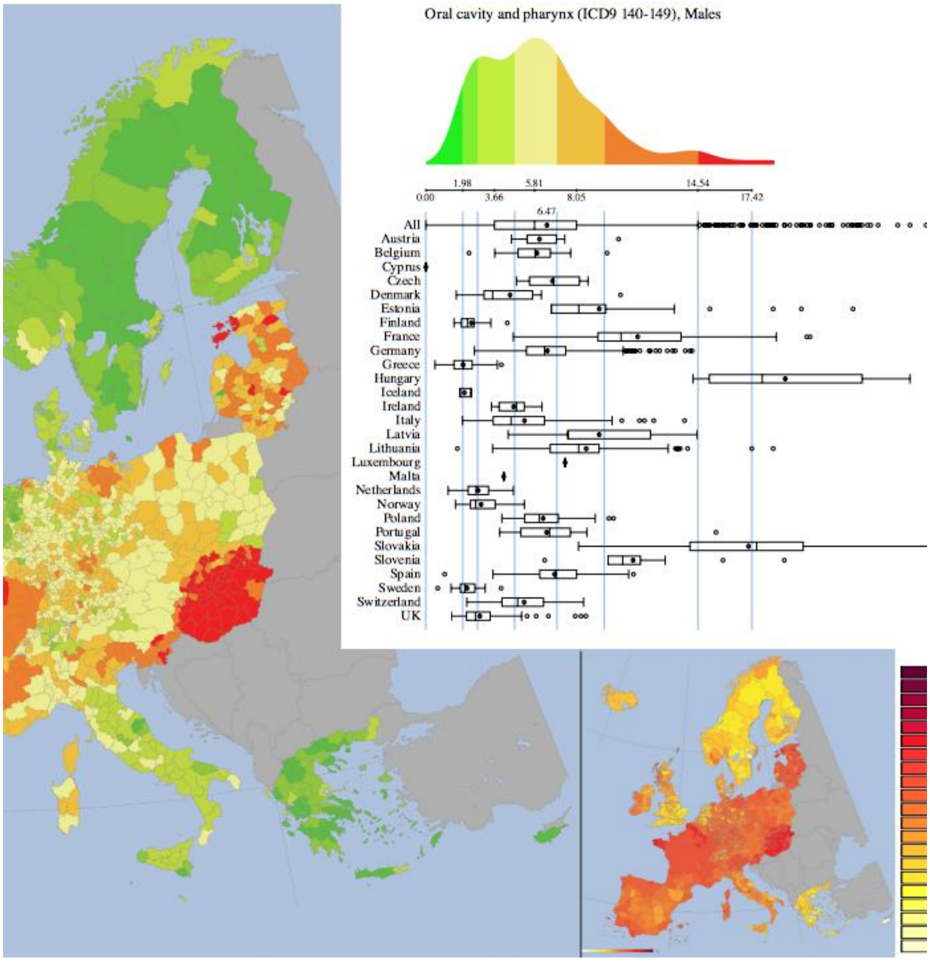


Figure 1. Images from Aerts, Clarke, and Keuper (2003) revealing two different classification schemes.

Source: Aerts, Clarke and Keuper (2003)

In Figure 3.6, Aerts, Clarke and Keuper (2003) explored the use of transparency (right) and colour (left) as an indicator of uncertainty in predicted urban growth.

Figure 3.7: Map pairs, boxplots and a probability distribution



Source: http://www.iarc.fr/en/publications/pdfs-online/epi/sp159/AtlasCancerMortalityEU-10.pdf

Figure 3.7, shows an example of map pairs, boxplots and a probability distribution. The large map shows Relative Risk of cancer incidence in Europe, while the small mirrored map shows the standard deviation of the relative risk.

Figure 3.8: Interquartile range

Indicator	Period	Rate	Lowest	Selected County Cancer Profile (Major sites)	Highest
All Sites	2011	423.6	332.5		527.6
Female Breast	2006-2011	111.4	80.1		143.7
Cervix	1996-2011	~	5.8		20.8
Colon and Rectum	2006-2011	42.0	39.5		70.3
Corpus and Uterus, NOS	1996-2011	~	12.4		37.2
Lung and Bronchus	2009-2011	80.8	50.9		115.4

Putnam
Statistically significant difference from the state rate: higher lower no difference
Missouri state average U.S. Combined (2010)
Quartile 1 Quartiles 2-3 Quartile 4

http://mcriaweb.col.missouri.edu/IAS/dataviews/report?reportId=13&viewId=3&geoReportId=62&geoId=1&geoSubsetId=

4. Report: Communication design workshop for the National Cancer Atlas

In September 2015, a collaborative Uncertainty Communication Design workshop was conducted with the project partners of the National Cancer Atlas project. These partners included representatives of the Cancer Council QLD, the National Health Performance Authority and the Queensland University of Technology.

The aim of the workshop was to scope the potential audiences of the National Cancer Atlas, explore key messages that should be communicated in the National Cancer Atlas and discuss the concepts about uncertainty that would be specific to these different audiences.

Attendees

Kerrie Mengersen (QUT)
Peter Baade (CCQ)
Joanne Aitken (CCQ)
Fiona Harden (QUT)
Jessie Roberts (QUT)
Susanna Cramb (CCQ
William Watson (NHPA)
Tomasz Bednarz (QUT/CSIRO)

Topics explored within the workshop

1. Why is communicating uncertainty an important problem?
2. Who are the Audiences of the National Cancer Atlas and what are their characteristics?
3. Can these audiences be grouped by the level of information detail they require?
4. What will the Atlas report (output measure or measures)?
5. What are the sources of uncertainty within the Atlas?

4.1 Why is Communicating Uncertainty Important

Workshop participants were asked to consider why communicating uncertainty is an important problem. The resultant discussion was framed in three different contexts with increasing focus. First, within science communication generally. Second, within the field of geospatial health statistics or disease mapping. Finally, within the specific context of the National Cancer Atlas.

Results of this discussion are summarised below.

Why is communicating uncertainty important:

1. **In Science & Science Communication Generally?**

   Uncertainty was recognized as being an important tool when evaluating the quality of scientific research and to assess the reliability of data driven insights.

An understanding of the uncertainty of statistical estimates is important to facilitate comparisons of research methods and comparing results between similar studies.

Higher levels of uncertainty around a measure indicate less is known about that measure. Therefore incorporating uncertainty, and highlighting areas of large uncertainty provides critical guidance when considering future research priorities by focusing on gaps in our current knowledge.

Many analytical and statistical methods can generate estimates, however unless we have comparable estimates of uncertainty around those measures it is impossible to appropriately evaluating and compare the effectiveness and reliability of new methods.

Appropriate measures of uncertainty supports valid interpretations of the estimates generated by a study and therefore leads to appropriate applications of insights from these results to real world settings.

Important in evaluating the accuracy and performance of new technology and methodologies.

Lack of uncertainty provides an unrealistic assurance that the results from research studies are facts rather than uncertain estimates. When the uncertainty about these estimates is uncovered by new scientific discoveries, it can degrade the general public's trust in scientific insights and degrade the reputation of science generally. The public does not know which estimate to place most reliance on or who to trust for information. By being transparent about the level of uncertainty in estimates, it makes it less surprising when contrasting results become available. Correspondingly, hiding uncertainty hides the inaccuracies present in all scientific discoveries and present the scientific process to be more solid than it actually is.

Along these same lines, the clear communication of uncertainty would better inform the general audience of the scientific process and the role of uncertainty in the calculation and reporting of scientific results.

Incorporating uncertainty enables us to assess changes in uncertainty over time. For example, if uncertainty in an estimate is reducing over time, then this would be a demonstration of an improvement in knowledge and/or methods over time.

2. **In Geospatial Data and Disease Mapping**
    a. Creating a map of modelled disease occurrence or risk can present estimates as more certain and accurate than they may actually be. This can be particularly misleading and lead to suboptimal decision making.
    b. Data aggregation decisions can influence final model outputs. Uncertainty can be useful tool in both evaluating which decisions lead to the most accurate results and also can make these inaccuracies more transparent.

3. **In the National Cancer Atlas**
    a. The small sample sizes present in some of the regions within the Cancer Atlas can lead to uncertain estimates. while the model outputs may be our best estimate, it is important that decision makers understand when the current ‘best’ is still uncertain.
    b. Provides a guide to applying these insights to policy developments. Informs decisions makers regarding the accuracy and reliability of estimates
    c. Uncertainty is important in applying the regional generalisations from the atlas to individual situations.
    d. Uncertainties support the need for future research in cancer outcomes and can help prioritize research projects.
    e. inclusion of uncertainty enhances the research output of the atlas. - tells the whole story and communicates clearly our current state of knowledge about inequalities in cancer incidence and survival in Australia
    f. provides examples of uncertainty communication methodologies for other Cancer Councils.

4.2 Who are the Audiences of the Atlas and what are their characteristics?

Through previous discussions with the NHPA and CCQ, eight target audiences had been identified for the National Cancer Atlas. Within the workshop participants defined the characteristics of each of these audiences, in terms of: The key messages we want to communicate to them through the Atlas? What decisions or questions is each audience trying to answer when exploring the Atlas? what is the skill levels of each audience (in terms of formal statistical training and analytical skills)? what potential risks are there when including uncertainty in the Atlas (misinterpretations, disregard info, etc)? what potential benefits could each audience gain from including uncertainty in the Atlas? what level of interest does each audience have in regards to the uncertainty of the map?

The following section details the 8 different audience profiles developed within the workshop.

Audiences:

1. General Audience/ General Public
2. Media
3. Government, lobby groups and health policy makers and advisors
4. Health managers
5. Regional
6. Local
7. Clinicians
8. Cancer patients and their carers, family or supporters
9. Researchers
10. Other Cancer Councils and Health Reporting organisations

4.2.1 General Audience

Key Messages:

- highlight the regions with variation in cancer incidence and survival.
- Show any relationship between cancer risk/survival and socio-demographic or rurality variables.

Knowledge & Skills:

- Formal Statistical training: low
- Analytical skills: Low

Decisions or Questions:

- How does my region compare to other regions in Australia.
- What are the reasons for areas of low or high risk?
- have I ever lived in an area of high risk?

Interest in uncertainty?

- minimal - most probably not aware of the presence of uncertainty.

Risks of including uncertainty:

- Key messages could be lost in information overload.
- too complex/difficult graph to interpret . There is a risk the audience will disengage.

Benefits of including uncertainty

- May calm an over-reaction to high risk regions.

4.2.2 Media

Key Messages

- Simple, short, graphs, infographics that are accurate and sharable.
- Where is there the greatest variation in cancer outcomes geographically. Are there any reasons why these areas have greater variation?
- "this is really important work"
- "this is innovative work"
- clearly explain the uncertainty in any high risk regions. Provide examples and words they can use to embed the uncertainty into their media messaging.

Skills

- Formal Stats training: low
- Analytical skills: low to medium
- May have some specialist training

Decisions & Questions

- Looking for a hook
- Is this newsworthy?

- Where are the highest risks?
- What is the government doing about these inequalities in cancer outcomes?
- What resources are available for people most at risk or with the highest needs?

Interest in Uncertainty
- averse to uncertainty
- Confuses the hook. looking for simple, clear news stories.

Risks of including uncertainty
- Misinterpretation or misrepresentation?
- May misinterpret uncertainty for poor quality research?

Benefits of including uncertainty
- General promotion and education about uncertainty.
- may reduce anxiety in small regions with high risk incidence. e.g. "A high risk in Mackay doesn't mean that everyone in Mackay will get cancer. "

Notes:
- need to provide sharable grabs, images, visualisations of infographics.

4.2.3 Government, lobby groups, health policy makers and advisors

Key Messages
- Are the current cancer treatment, screening and support services sufficient?
- Are there inequalities and if so, where?
- Are the government programs working? Jas there been a change over time?
- How does their jurisdiction compare to others?
- What are the highest priority interventions and research for the future.
- So what -> how best to translate these insights into policy.
- What are the most pressing inequalities in cancer outcomes and are there any recommendations for addressing these?

Skills
- Formal Stats Training: low to medium
- Analytical skills: medium to high
- Other: mostly communication and decision making skills not statistical

Decisions and Questions
- What can we do to improve survival rates and reduce inequalities in cancer outcomes?
- Can we show our current or recent health services are creating change?

Interest in uncertainty
- uncertainty can be confusing and can hinder or slow down decision making. Often viewed as a bad thing and decision makers would generally want to see a definite number.

- May not know how to apply the uncertainty in their current decision making framework.
- Could be seen as a valuable tool if presented the right way or if they have had sufficient training.

Risks of including uncertainty

- Information may be regarded as of poorer quality if uncertainty but of greater long term benefit because policy will be developed for better future outcomes.
- May lead to decision paralysis
-

Benefits of including uncertainty

- Better represents our current state of knowledge
- Uncertainty may help quantify how much money should be spend on a program and when. May be very valuable in designing milestones and clarification points for a health program. May mean policy decisions are made that embed flexibility when the current state of knowledge contains uncertainty.

Other Notes

- Need to ensure that the scientific evidence provided can inform the decision making process.
- This audience will have many competing priorities.
- Likes to be able to show improvement over time.

4.2.4 Cancer Patients/Survivors and their family, carers and friends

Key Messages

- Insights at their community.
- Where can they access services, support and information

Decisions and Questions

- What are the benefits of my different treatment options and ancillary side effects?
- What is the best treatment available to me?
- What have other people with my cancer diagnosis, and/or in my region,s done? What services did they access? What treatment did they have?
- Is the risk of survival lower or higher in my region?
- what treatment options are available to be in my regions?
- What resources are available in my region?
- How far do I have to travel for my treatment?
- What resources are available to me in my community?
- Is there a lower than average risk of survival in my region? If so why? What can I do about it?
- How does my community compare to other similar communities? (in the same peer group)
- Looking for more accurate information to replace "Dr Google"

Skills and Knowledge

- Formal statistical training: overall low, but highly varied
- Analytical skills: overall low, but highly varied
- Will be looking for more accurate information to replace Dr Google.

Interest in uncertainty

- How likely is my treatment to be unsuccessful/ successful
- could provide comfort for people living in high risk areas.
- May affect their life and their treatment choices.
- uncertainty about life and treatment options will lead to high anxiety.

Risks of including uncertainty in the Atlas

- The uncertainty may create greater physical and emotional stress for the patient and their family. Difficulty of the unknown and not having a clear right answer.

Benefits of including uncertainty in the Atlas

- may enable more informed decisions about how they manage their treatment.
- may provide comfort if they live in an area of high risk. (for example if their family also live in the same region)

4.2.5 Researchers

Key Messages

- Here are the gaps in our knowledge.
- Here is the uncertainty in our outputs.
- The methods we used for developing these disease maps are accurate and robust.
- The methods we used to communicate the uncertainty are clear and accurate.
- Our methods of communicating /representing uncertainty have been successful and are accessible to non-expert audiences and decision makers.
- our research is awesome and our methods robust. !!!

Decisions and Questions

- What is the quality/accuracy/uncertainty of the estimates?
- Are the inferences made from the data appropriate.
- What is the current state of knowledge in this area, current best practice?
- What are the gaps in the current knowledge, how can this research relate to my research
- Are these methods applicable to my area of research?

Skills & Knowledge

- Formal statistical training: high
- Analytical skills: high

What does uncertainty mean to this audience ?

- highlights the quality of the research
- highlights where future research should focus

- guides the application of the scientific insights to real world practice

Interest in uncertainty

- High

Risk of including uncertainty information

- minimal

Risks of excluding uncertainty information

- excluding uncertainty information can give a false representation of our current state of knowledge. This could result in important research problems of knowledge gaps being missed because our knowledge us presented more certain than it is.
- Inaccuracies are missed and future research is misguided.
- missed opportunities for research and for patient outcomes.

Benefits of including uncertainty information

- Clear spotlight on future research opportunities.
- Clear support for the need of research they may be applying for funding for.

4.2.6 Health Managers (Regional and Local)

Key Messages we wish to convey to this audience

- Where are the demands for services greatest?
- These regions need to focus on these support services...
- Quantify what the needs of their region are.
- These are the services available in your region

Decisions and Questions

- How do I budget and allocate resources to best meet the needs of residents in my region.
- How does my region /jurisdiction compare to other regions in Australia? Better/worse/same..
- What services are available in my region and what services should I be advocating for?
- Are there any shortfalls in screening or support services in my regions?
- Do I need to budget any extra services to meet the needs of this group?
- Are these results what I expect? better/worse/the same?

Skills

- Formal statistical training: medium
- Analytical skills: medium

Interest in uncertainty

- low to medium

Risks of including uncertainty in communications

- Confusing or difficult to understand (time poor audience)

Risks of excluding uncertainty information

- State of knowledge and information appear more accurate than they actually are.
- Recommendations and advice to patients could be represented as more solid than it actually is
- high or low risk in their region may be interpreted as more certain or accurate than it actually is. Leading to over/under prescription.

Benefits of including uncertainty in communications

- Helps ensure that health strategies and spending are meeting real needs
- Optimise cashflow (reduce the risk of spending money when the estimates/insights are not reliable)

4.2.7 Clinicians

(Similar to health managers)

Key Messages

- Information on the needs of the region they work in.
- Type of services available, and should be provided to this patient group.
- which Regions have higher than average risk of cancer incidence or lower survival
- which Regions that have higher needs or are a higher 'disadvantage' (due to rurality or socio-demographic aspects)

Skills & Knowledge

- Formal statistical training: low
- Analytical skills: low

Decisions or Questions

- What services do I need to ensure are available in the region I work .
- Do I need to promote a higher rate of screening in my region?
- Do I need to promote the services that are available in my region? (e.g support for travel, or other support, treatment options that might be impacted by travel challenges)
- Are residents in my region facing greater challenges due to socio-economic or geographic boundaries?

Interest in Uncertainty

- low to medium - time poor.

Risk of including uncertainty information

- May overwhelm a time poor audience. They may give up on the atlas because the uncertianty makes it difficult to digest the information quickly.

Benefits of including uncertainty information

- Can clarify the Atlas outputs and ensure that under or over treatment is not prescribed due to estimates appearing more accurate or certain than they actually are.

4.2.8 Other Cancer Councils and Health Reporting Organisations

Key Messages

- Information about the areas of inequality that need to be addressed.

Decisions and Questions

- Where are the greatest needs for intervention?

Skills

- Some staff would have formal Statistical training to guide internal interpretation
- Analytical Skills

Risks of including uncertainty in Atlas Communications

- Key messages are not as easy to communicate.

Potential benefits of including uncertainty in Atlas Communications

- Help guide and prioritise future research

4.3 Organising Audiences by the level of information complexity they require

Workshop participants considered four levels of increasing information complexity and considered for each level; the information complexity or detail within that level, the most appropriate audience and potential communication products.
The outcomes of the discussion are summarised in Table 4.
Insert a nicer graphic

Table 4.1; Summary of discussion outcomes

		Audience	Potential product
1	**executive summary** (short clear statements of insights)	media	**??**
2	**Map + results & uncertainty** (results and uncertainty information presented in a formal accessible to a non-expert)	media, general audience, cancer patients & carers	**??**
3	**Map + numbers + technical measures of uncertainty** (includes technical estimates and uncertainty)	cancer patients & carers clinicians ____ Health policy advisors ___ health managers	**??**
4	**Technical report + data set + ?** (Contains details of methods, access to data, other statistical outputs)	Researchers	**??**

4.4 What will the Atlas Report?

The group discussed the different report measures used cancer mapping and what may be most appropriate for the National Cancer Atlas. This discussion highlighted the wide variety of potential measures that can be used for cancer mapping, and the lack of accepted standards. There was a lack of clarity among the group of the merits and disadvantages of the different output measures and this was identified as an area of further research as well as an area of potential linguistic imprecision.

Points to consider when deciding on a report measure:

- Media needs to be able to grasp outputs quick and easily. Outputs from the QLD cancer Atlas aren't intuitive for the media team.
- Probability is easily all ready understood by media and the general public
- Is it better to report positive rather than negative, survival vs death, etc?
- Consider contexualising probabilities – x number in 100, etc.
- Further discussion required to decide which measure to report. potentially there will be more than one.
- How does interpretation differ for regional estimates and individuals. How do we help users apply the regional population estimates to an individual?

4.4 Sources of Uncertainty

Workshop participants split into groups and discussed the sources of uncertainty important for each of the identified audiences for the National Cancer Atlas.

Table 4.2: Sources of uncertainty

Data		**Model**		**Outputs**
	Methodologies	**Model Assumptions**	**Method disagreements**	**linguistic uncertainty**
- Estimated population of each regions (ABS) **- Estimated demographic breakdown of each region (ABS)** **- Socio-economic status is generalised across the entire region.** **- Classification uncertainty around the cause of death.**	- Smoothing algorithm - Model prior distributions (may also be a input rather than a method) - ???	- residential address does not contain any info of length of time at that residence. - ???	- smoothing algorithm - ???	- meaning of: probability, uncertainty, risk, cause, correlation, random - ???

- Classification uncertainty around indigenous identification **- Residential address does not contain any info of time at that residence or region.** **- ???**				

4.5 Further Discussions

Throughout the workshop we identified several questions important to the cancer atlas that require a more in depth discussion than was possible within the workshop. These discussion points are listed below.

1. Reliability vs Confidence Vs Certainty vs Uncertainty

- Uncertainty has different meanings to different audiences of the atlas. For Government and policy-makers, as well as health care managers it is about reliability or confidence. For researchers, uncertainty can represent opportunity.

2. Emotional response of audience and focusing on positive stories rather than negative

3. Embedding cancer stories within the map

4. What does normal variation in Cancer risk and survival look like geographically?

References

Aerts, J. C. J. H., Clarke, K. C. and Keuper, A. D. (2003). Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science,* 30 (3), pp. 249–261.

Bedford T., & Cooke R. (2001). Probabilistic risk analysis: foundations and methods. *Cambridge University Press*. Cambridge, UK.

Begg, S. H., Bratvold, R. B., & Welsh, M. B. (2014, May). Uncertainty vs. Variability: What's the Difference and Why is it Important?. In *SPE Hydrocarbon Economics and Evaluation Symposium*. Society of Petroleum Engineers.

Bertin, J. (1981). *Graphics and graphic information processing*. Walter de Gruyter.

Brodlie, K., Osorio, R. A., & Lopes, A. (2012). A review of uncertainty in data visualization. In *Expanding the frontiers of visual analytics and visualization* (pp. 81-109). Springer London.

K. Brodlie, R.A. Osorio, and A. Lopes (2011). A Review of Uncertainty in Data Visualization, *Expanding the Frontiers of Visual Analytics and Visualization, part 2,* J. Dill et al., eds., Springer, pp. 81−109.

Brown, R. (2004). Animated visual vibrations as an uncertainty visualisation technique. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia - GRAPHITE '04* (Vol. 1, pp. 84–89). ACM Press. http://doi.org/10.1145/988834.988849

Cedilnik, A, Rheingans, P (2000). Procedural Annotation of Uncertain Information: *Proceedings of IEEE Visualization IEEE 2000*, 77–84. URL: http://www.cs.umbc.edu/~rheingan/pubs/grids00.pdf

Deitrick, S., & Edsall, R. (2008). *Geographic Visualization*. (M. Dodge, M. McDerby, & M. Turner, Eds.), *Geographic Visualization: Concepts, Tools and Applications*. Chichester, UK: John Wiley & Sons, Ltd. http://doi.org/10.1002/9780470987643

Stephanie Deitrick & Elizabeth A. Wentz (2015) Developing Implicit Uncertainty Visualization Methods Motivated by Theories in Decision Science, Annals of the Association of American Geographers, 105:3, 531-551, DOI: 10.1080/00045608.2015.1012635

Enserink, B., Kwakkel, J. H., & Veenman, S. (2013). Coping with uncertainty in climate policy making: (Mis)understanding scenario studies. *Futures*, *53*, 1–12. http://doi.org/10.1016/j.futures.2013.09.006

Evans, B. J. (1997). Dynamic display of spatial data-reliability: does it benefit the map user? Computers and Geosciences 24(1): 1–14.

Flavell, J. H. (1976). "Metacognitive Aspects of Problem Solving". In L. Resnick (Ed.). The Nature of Intelligence (pp.231-236). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Funtowicz, Silvio O., and Jerome R. Ravetz (1985). Three types of risk assessment: a methodological analysis. In *Risk Analysis in the Private Sector, ed. C.* Whipple and V. T. Covello, 217-31. New York: Plenum.

Gershon ND (1992) Visualization of fuzzy data using generalized animation. *Proceedings of Visualization 92*, Boston, Massachusetts, IEEE Computer Society, Los Alamitos, California, pp 268±273

Grubler, A., Ermoliev, Y., & Kryazhimskiy, A. (2015). Coping with uncertainties-examples of modeling approaches at IIASA. *Technological Forecasting and Social Change*, *98*, 213–222. http://doi.org/10.1016/j.techfore.2015.06.004

Griethe, H. & Schumann, H. (2006). The Visualization of Uncertain Data: Methods and Problems, In *Proceedings of SimVis 2006*, 143–156, Magdeburg, Germany.

Han, P. K., Klein, W. M., & Arora, N. K. (2011). Varieties of Uncertainty in Health Care A Conceptual Taxonomy. *Medical Decision Making*, *31*(6), 828-838.

Harrower, M., & Street, N. P. (2003). Representing Uncertainty : Does it Help People Make Better Decisions ?, (1)

Hughes, R. I. G. (1989). The structure and interpretation of quantum mechanics. Harvard University Press, Cambridge, Massachusetts, USA.

Hunter, G. J., & Goodchild, M. F. (1996). Communicating uncertainty in spatial databases. *Transactions in GIS*, *1*(1), 13–24. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0029731784&partnerID=tZOtx3y1

Johnson, C. (2004). Top scientific visualization research problems. *IEEE Computer Graphics and Applications*, *July/August*, 13–17.

Johnson, C. R., & Sanderson, A. R. (2003). A next step: visualizing errors and uncertainties. *IEEE Computer Graphics and Applications*, 6–10.

Kahneman, D., & Tversky, a. (1982). Variants of uncertainty. *Cognition*, 11(2), 143–157. doi:10.1016/0010-0277(82)90023-3

Kinkeldey, C., MacEachren, a M., & Schiewe, J. (2014). How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *Cartographic Journal*, *51*(4), 372–386. http://doi.org/10.1179/1743277414Y.0000000099

Knight, F. H. (1921). Risk, uncertainty and profit. New York: Houghton Mifflin Company (republished in 2006 by Dover Publications, Mineola, NY).

Kubicek, P., Sasinka, C. (2011). Thematic uncertainty visualization usability- Comparison of basic methods. *Ann. GIS*, *17*, 253–263.

Kujala, H., Burgman, M. A., & Moilanen, A. (2013). Treatment of uncertainty in conservation under climate change. *Conservation Letters*, *6*(2), 73–85. http://doi.org/10.1111/j.1755-263X.2012.00299

Lindley D. (2006) Understanding uncertainty. Wiley-Blackwell.

Lipshitz, R., & Strauss, O. (1997). Coping with Uncertainty: A Naturalistic Decision-Making Analysis. *Organizational Behavior and Human Decision Processes*, *69*(2), 149–163. http://doi.org/10.1006/obhd.1997.2679

MacEachren, A. M. (1992). Visualizing uncertain information. *Cartographic Perspectives*, (13), 10-19.

MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., & Hetzler, E. (2005). Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, *32*(3), 139–160. http://doi.org/10.1559/1523040054738936

Morgan, M. G., & Henrion, M. (1990). Uncertainty: a Guide to dealing with uncertainty in quantitative risk and policy analysis Cambridge University Press.New York, *New York, USA.*

Matthews, M., Rehak, L., Famewo, J., Taylor, T. and Robson, J. (2008), Evaluation of New Visualization Approaches for Representing Uncertainty in the Recognized Maritime Picture, (DRDC Atlantic CR 2008-177) Defence R&D Canada - Atlantic, Dartmouth, NS (CAN); Humansystems Inc., Guelph ON (CAN).

Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization. The Visual Computer, 13(8), 370–390. doi:10.1007/s003710050111

Pang, A.T., Wittenbrink, C.M. and Lodha, S.K. (1997), Approaches to uncertainty visualization, *The Visual Computer*, 13 (8), 370-390.

Potter, K., Rosen, P., & Johnson, C. R. (2012). From Quantification to Visualization: A Taxonomy of Uncertainty Visualization Approaches. In *Uncertainty Quantification in Scientific Computing IFIP Advances in Information and Communication TechnologyVolume 377, 2012,* (pp. 226–249).

Politi, M. C., Clark, M. A., Ombao, H., Dizon, D., & Elwyn, G. (2011). Communicating uncertainty can lead to less decision satisfaction: a necessary cost of involving patients in shared decision making?. *Health Expectations*, *14*(1), 84-91.

Regan, H. M., Colyvan, M., Burgman, M. a, Applications, E., & Apr, N. (2008). A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology, 12(2), 618–628.

Regan, H. M., Colyvan, M., & Burgman, M. a. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications*, *12*(2), 618–628. http://doi.org/10.1890/1051-0761(2002)012[0618:ATATOU]2.0.CO;2

Sanyal, J., Zhang, S., Bhattacharya, G., Amburn, P., & Moorhead, R. J. (2009). A user study to compare four uncertainty visualization methods for 1D and 2D datasets. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 1209–18. http://doi.org/10.1109/TVCG.2009.114

Slocum, T. A., Cliburn, D. C., Feddema, J. J., & Miller, J. R. (2003). Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science*,*30*(4), 299-317.

Thomson, J., Hetzler, E., Maceachren, A., & Gahegan, M. (2005). A Typology for Visualizing Uncertainty. Visualization and Data Analysis, 5669(January), 146–157. doi:10.1117/12.587254

Thunnissen DP. Uncertainty classification for the design and development of complex systems. In: Proceedings of the third annual predictive methods conference— PMC2003, Newport Beack, California, USA; 16–17 June 2003

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment, 4(1), 5–17. doi:10.1076/iaij.4.1.5.16466

Thunnissen DP. Uncertainty classification for the design and development of complex systems. In: Proceedings of the third annual predictive methods conference— PMC2003, Newport Beack, California, USA; 16–17 June 2003

Uusitalo, L., Lehikoinen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support.*Environmental Modelling & Software*, *63*, 24-31.

Viard, T., Caumon, G., Lévy, B. (2011). Adjacent versus coincident representations of geospatial uncertainty. Which promote better decisions? *Computer. Geosci*, *37*, 511–520.

Wilkinson L. (1999). The Grammar of Graphics. Springer-Verlag New York, Inc.

Vullings, L., Blok, C., Wessels, C., Bulens, J. Dealing with the Uncertainty of Having Incomplete Sources of Geo-Information in Spatial Planning. *Applied Spatial Analysis and Policy* 1-21 (2013)

Wittenbrink, C. M., Pang, A. T., & Lodha, S. K. (1996). Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics*, *2*(3), 266–279. http://doi.org/10.1109/2945.537309

Zuk, T. and Carpendale, S. (2007), Visualization of Uncertainty and Reasoning, In *Proceedings of Proceedings of the 8th international symposium on Smart Graphics*, 164-177, Kyoto, Japan.

Appendices

Appendix A: Sources of Uncertainty

Framework I

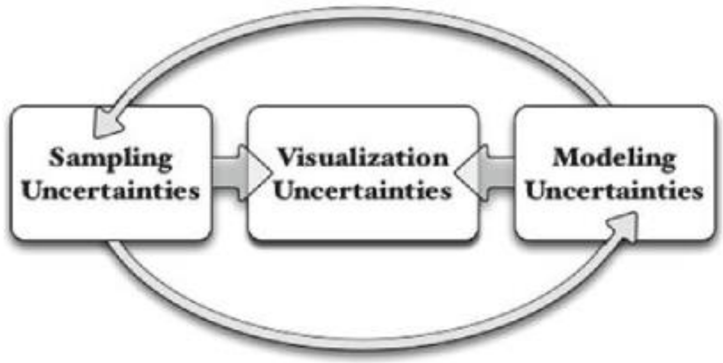


Fig. 1.1 Sources of uncertainty. Both sampling and modeling uncertainties affect each other and add to visualization uncertainties

Sources of Uncertainty within Modeled Outputs

Sophisticated computational models often contain elements designed to estimate the uncertainty or variability in the model predictions. Sources of this type of uncertainty include:

- residual variability from simplifying abstractions.
- variability in the mechanism or magnitude of causality and relationships
- potential error in model inputs
- incorrect model parameters
- imprecision in the tacit knowledge incorporated in the model.

Framework II

There have been several attempts to create taxonomies of different sources of uncertainty. I have found the characterisation of sources provided by Morgan and Henrion (1990) to be the most useful and generally applicable to any scenario where outputs from statistical models are being used to inform decision making. Morgan and Henrion consider that the most appropriate method to characterize uncertainty, and the most appropriate method for trying to reduce it, generally depends on the particular kind of source. Hence, they provide the following list of different kinds of sources from which uncertainty can arise:

- Statistical variation and random error
- Subjective judgment and systematic error
- linguistic imprecision
- variability
- inherent randomness
- disagreement
- approximation

Morgan and Henrion (1990) provide a description for each of these sources.