The epidemiology, healthcare, and societal burden of basal cell carcinoma in Wales 2000-2018: A retrospective nationwide analysis

Nader Ibrahim,[1,2] Matthew Jovic,[1,3,4] Stephen Ali,[1,2] Namor Williams,[5] John A.G. Gibson,[1,2] Rowena Griffiths,[3,4] Thomas D. Dobbs,[1,2] Ashley Akbari,[3,4,6] Ronan A. Lyons,[3,4,6] Hayley A. Hutchings[3] and Iain S. Whitaker[1,2]

[1]Reconstructive Surgery & Regenerative Medicine Research Centre (ReconRegen) Swansea University Medical School, Institute of Life Sciences, Swansea, SA2 8PP, UK

[2]The Welsh Centre for Burns & Plastic Surgery, Morriston Hospital, Swansea, SA6 6NL, UK

[3] Swansea University Medical School, Institute of Life Sciences, Swansea, SA2 8PP, UK

[4] Population Data Science, Health Data Research UK, Swansea University, Swansea, SA2 8PP, UK

[5] Department of Pathology, Singleton Hospital, Swansea Bay University Health board, Swansea SA2 8QA, UK

[6]Administrative Data Research Wales, Swansea University, Swansea, SA2 8PP, UK.

Corresponding Author: Mr Nader Ibrahim, MRCS, BSc (Hons), MBBS
Email: nader.ibrahim@swansea.ac.uk
ORCiD: https://orcid.org/0000-0003-4609-1430

Abstract

Background:
Basal Cell Carcinoma (BCC) represents the most commonly occurring cancer worldwide within the Caucasian population. Reports predict 298 308 cases of BCC in the UK by 2025, at a cost of £265-366 million to the National Health Service (NHS). Despite the morbidity, societal and healthcare pressures that manifest, routinely collected healthcare data and global registration remains limited.

Objectives:
To calculate the incidence of BCC in Wales between 2000-2018 and to establish the healthcare utilisation and estimated cost of care.

Methods:
The Secure Anonymised Information Linkage (SAIL) Databank is one of the largest and most robust health and social care data repositories in the UK. Cancer registry data was linked to routinely collected healthcare databases between 2000-2018. Pathological data from Swansea Bay University Health Board (SBUHB) was used for internal validation.

Results:
61,404 histologically proven BCC were identified during the study period within the SAIL databank. The European age standardised incidence (EASR) for BCC in 2018 was 224.6 per 100,000. Based on validated regional data a 45% greater incidence was noted within SBUHB pathology versus matched regions within SAIL between 2016-2018. A negative association between deprivation and incidence was noted with a higher incidence in the least socially deprived and rural dwellers. Approximately 2% travelled 25-50 miles for dermatological services in comparison to 37% for plastic surgery. Estimated NHS costs of surgically managed lesions 2002-2019 equated to £119.2-164.4 million.

Conclusion:
Robust epidemiological data which are internationally comparable, and representative is scarce within non-melanoma skin cancer despite cases of individual improvement. The rising global incidence coupled with struggling healthcare systems in the post Covid-19 recovery period serves to intensify the societal and healthcare impact. This study is the first to demonstrate the incidence of BCC in Wales and one of a small number in the UK using internally validated large cohort datasets, and further demonstrates one of the highest published incidences within the UK and Europe. In the modern era of health informatics and advanced analytics it is imperative that we capitalise on routinely collected healthcare data. Therefore, it must be accurate, comprehensive, and accessible otherwise services will under-deliver.

Introduction

Basal cell carcinoma (BCC) is the most common malignancy within the Caucasian population worldwide, with an incidence rate that continues to rise[1]. While BCCs have a low mortality, they are associated with a significant degree of physical and psychological morbidity, as well as societal and healthcare costs. Despite this significant burden, BCC primary data collection are relatively poor in comparison to other common malignancies, and negatively impedes research and policy making. Routinely collected electronic health record (EHR) data and registration within many cancer registries remains less than fully representative, largely due to the associated workload, synchronicity and metachronicity of lesions, and non-integration of clinically diagnosed (but not histologically confirmed) and treated lesions[2]. The resulting impact of this current data collection strategy are a underestimation of the true burden of disease[3–8]. UK registries also only consider the first diagnosed BCC per patient as stipulated by the United Kingdom and Ireland Association of cancer registries (UKIACR), meaning that the true workload associated with the many patients who present with multiple lesions at once or in succession are underestimated. This practice is, however, not isolated to the UK[1].

In order to combat poor data collection novel methods have been used. These suggest an annual UK growth in incidence of 5%[7]. Historically the epidemiology of BCC in Wales has been poorly represented in the literature in a country in which 96% of its residents identify as Caucasian with a high proportion of Fitzpatrick skin types 1 and 2[9]. Regional population studies which have been carried out in Wales, although small, have highlighted a growth in registration of 66% in a 10-year interval[10].

These rising rates are predicted to result in 380,000 cases of NMSC (BCCs forming 80%) in the UK by 2025, with a projected economic impact to the NHS of £338-465 million[11].

The purpose of this study is to use population-scale data to curate an electronic-cohort (e-cohort) of anonymised healthcare data to determine the epidemiology, utilisation, societal and financial implications of BCC care across Wales over a period of 18 years. This represents the largest epidemiological study of its kind reinforced by validated pathological data and advanced data linkage to cross interrogate various data sources in Wales.

Methods

Study design and data sources

This population-based retrospective cohort study was conducted in accordance with the REporting of studies Conducted using Observational Routinely-collected Data (RECORD)[12] statement.

The Secure Anonymised Information Linkage (SAIL) Databank, a world leading trusted research environment (TRE) with population-scale anonymised healthcare data in Wales[13,14] was the primary source, supplemented by robust pathological data from Swansea Bay University Health Board (SBUHB).

The SAIL Databank holds billions of de-identified person-based records from multiple data sources, integrated through privacy protecting data linkage[15]. Routinely collected EHR are assigned a unique identifier referred to as an anonymised linking field (ALF), generated by Digital

Health and Care Wales (DHCW), which undergoes 2 further encryptions prior to project access within the SAIL Databank[15].

The e-cohort was defined within the Welsh Cancer Intelligence Surveillance Unit (WCISU) and linked with primary care data covering the vast majority of the general practice population, and secondary care inpatient and outpatient data covering 100% of the population (Table 1). These data were linked between the years 2000-2018 and were analysed to investigate the epidemiology, healthcare utilisation and cost of BCCs.

Pitfalls within coding of NMSC[2] arise due to generic coding such as the International Classification of Disease 10th revision (ICD-10) C44 code. Therefore, the most reliable method of identification of a BCC, are those that are histologically proven as such within cancer registries and linked using the unique patient identifier to various other databases where coding is less specific.

Study Population

All patients with a diagnosis of BCC registered within WCISU were considered within the study between 2000-2018. An e-cohort of patients was defined using the International Classification of Disease 10 (ICD-10) C44 code and International Classification of Diseases for Oncology 02 (ICD-O2) morphology codes 8090-8095, 8097 and behavioural code 3 (invasive). In situ disease was not included. Analysis took place on the first registered BCC per patient during the 18-year study period. Table 1 summarises the data sources used within SAIL.

Outcomes

Patient Demographics

Demographics, lesion locality and morphology were identified at the histological lesion diagnosis date. The Welsh Index of Multiple Deprivation (WIMD, 2011 version) is used, which is a tool approved by Welsh government to assess socioeconomic status of small geographical regions, Lower-layer Super Output Areas (LSOA), of which there are 1,909 within Wales[16]. Each LSOA is assigned a quintile of socioeconomic status from one, the lowest, to five, the highest.

Incidence

Incidence is described as the Crude rate, European age standardised rate (EASR)[17,18] and World age standardised rate (WASR)[19] to permit comparison with worldwide data. Mid-Year population estimates for specified age groups and regions were obtained from Stats Wales published by Welsh Government[20].

Healthcare Utilisation

Primary care attendances were accessed via WLGP, secondary care outpatient appointments via OPDW and inpatient or day case procedures via PEDW. Duplicate records were removed within each data source to permit a single ALF per patient.

Access to Health

Centroid coordinates of each patient's LSOA were used to calculate geodesic distances travelled for specialist treatment. A distance formula based on the earth's radius was used to give an approximation of distances travelled as the crow flies.

Costing

The financial burden of BCC care on NHS Wales is given based on works by Vallejo-Torres et al using a bottom-up and top-down cost model (please refer to publication for specific methodology), and subsequently used by Goon et al as a framework to estimate cost projection for the UK in 2025[11,21]. The top-down approach combined health service utilisation data with unit cost of services data, whilst the bottom-up approach constructed a simplified patient pathway which was subsequently populated with probabilities based on published literature and associated unit cost of each element.

Data Validation

Pathological data from SBUHB was extracted from a prospective pathology database by a consultant histopathologist (N.W.) between 2003-2019 using SNOMED CT codes for BCC (see Supplements Table 5). Two districts reliably encompassed within the jurisdiction of SBUHB (Swansea and Neath Port Talbot) were isolated; 4 years of data were selected and validated (2014-2017). Initial validation of pathological data took place via a filtering algorithm to highlight any BCC excision or biopsy within 3 months of another excision or biopsy with a matching NHS number to avoid double counting of lesions (i.e., those undergoing a biopsy prior to excision or those which were incompletely excised and required further surgery). We opted to use 3 months as the upper boundary cut off as this represents the urgent suspected cancer pathway and referral to treatment timeframe[22]. Histopathology reports which were within 3 months of one another were then individually searched by a single plastic surgeon (N.I). Incorrectly coded tumours were eliminated, lesions biopsied and subsequently excised were counted once, incompletely excised lesions were counted once. Electronic documents were consulted if details from the histopathology report were insufficient. An adjustment factor was applied to all remaining pathological data based on the validation. Comparison in case burden was then made between adjusted pathology data and SAIL data for matched regions and years.

Statistical Analysis

SAIL data were extracted by a SAIL data analyst using SQL developer IBM DB2 (IBM Corp, Armonk, NY), Microsoft Excel 2016 (Microsoft Corp, Redmond, Washington USA) and R i386 Version 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria) were used for statistical analysis. Mathematica Version 12.3 (Wolfram Research, Inc., Champaign, Illinois, USA) was used for data forecasting. A statistical forecasting model was constructed within SPSS based on linear and quadratic equations to estimate 3 case burden scenarios extrapolated to 2030 for Wales.

Results

61,404 patients with a histologically proven BCC for the duration of the study period were identified between 2000-2018 in Wales. The population of Wales increased by 7.9% during the study period to 3,138,631.

Data acquisition from the WCISU is illustrated in Fig. 1 and population demographics are summarised in Table 2. The median age of the first histologically proven BCC for both males and females for the duration of the study period was 72 (Male 95% CI 70.5-70.7, Female 95% CI 70.3-70.6) with a male preponderance (1:0.83).

Incidence

As confirmed through direct correspondence with WCISU, data collection for non-melanoma skin cancer was made mandatory from 2016. This accounts for the large apparent increase in recorded incidence in 2016. The mean crude incidence from 2016-2018 was 242.5 per 100,000 person years (PY) (95% CI, 222.2-262.8), the EASR was 235.3 (95% CI, 209-261.5) and WASR was 117.2 per 100,000 PY (95% CI, 104.2-130.1) for the first registered BCC per patient. For the period 2016-2018 a mean of 7,579 cases were diagnosed per year. Inferences on incidence change are unlikely to be representative as data capture was incomplete pre 2016 as exemplified by Figure 2. Age specific incidence per decade age group between 2016-2018 is represented in Figure 3. All forms of standardisation methodology are represented in Table 6 within supplementary data.

Pre-2016 Projection

Prior to 2016, data collection for non-melanoma skin cancer was non-mandatory within WCISU, resulting in under-reporting of incidence. Comprehensive pathology data from SBUHB for the districts of Swansea and Neath Port Talbot districts have been consistently collected from 2003. Four years of this dataset were validated, and an adjustment factor applied to the remaining cohort. This data was compared directly to matched regions within the SAIL Databank and used to backward project the pre-2016 data to give a more accurate reflection of pan Wales incidence (Fig.4).

Body Site

The face was the most affected anatomical region in both males and females (34.9% in males and 39% females). BCCs of the face, lower limb, lip, and eyelid were more common in females and BCCs of the ear/external auditory canal, scalp/neck, trunk, and upper limb were more common in males (Table 2). Pearson's Chi Squared displayed a clear statistical significance between site and gender ($P<<0.05$).

Socio-economic and geographical distribution

Crude incidence based on urban/rural locality is demonstrated in Fig. 6. Despite the diagnostic caseload within urban regions being three times higher than that recorded for rural regions (Fig 5.); relative to the population, crude incidence is demonstrated to be higher within a rural locality. The mean crude incidence was statistically significant using One-Way Anova. The Tukey's Honest significant difference test was used to further delineate statistical significance between each of the three categories($P<<0.05$).

A greater proportion of the study group were found to be within the least socially deprived socioeconomic group (WIMD 5) in both males and females (Table 2). No statistical difference was noted between the median WIMD of males versus females using the Mood Median Test ($p>0.05$).

Healthcare Utilisation - Outpatient Appointments

16,016 records were identified as BCC related outpatient appointments between 2004-2018, 11,686 squamous cell carcinoma (SCC) and the remaining 14,306 are coded generically as C44 which encompasses all non-melanoma skin cancer. The mean number of follows ups for BCC coded outpatient appointments (based on linked morphology codes) was 2.42 (median 1, IQR 1-3). 84% of referrals were from GPs, 9% were referred from other consultants (excluding Accident and Emergency), 4% from other sources and 3% from the consultant responsible for the OPDW episode. The proportion of outpatient appointments by speciality are illustrated in Table 3.

Healthcare Utilisation - Inpatient/Daycase treatment

61,284 inpatient/daycase procedures were coded as BCC related and histologically proven as such, representing 39,803 patients. The mean spell duration was 0.5 days (median 0, IQR 0,0). The mean number of episodes per patient was 1.5 (median 1, IQR 1-2) during the study period. The cumulative number of bed days was 28,682. The distribution amongst inpatient/daycase speciality is outlined in Table 4 and distance travelled to obtain specialist treatment in Fig.8.

Geographical variation

The EASR of each principal region within Wales based on incidence data for 2016-2018 is represented in the choropleth map (Fig.7). The EASR was noted to be highest in urban Swansea (318.9 per 100,000 PY, 95% CI 296.1-341.6) with a population of 245,000, almost twice as high as that noted in rural Ceredigion (172.6 per 100,000 PY, 95% CI 144.8-200.5) which has a population of 73,000 (stated populations based on 2018).

Projecting future cases

Analysis of validated SBUHB pathology data shows a 45% greater incidence compared to matched regions within the SAIL Databank for the period 2016 to 2018 (based on Swansea and Neath port Talbot). Linear and quadratic projections of the adjusted and scaled SBUHB pathology data to all regions within Wales and adjusted SAIL data are depicted within Fig. 8. Illustrated are three scenarios for each dataset based on projections of growth (see supplements for further information regarding projection methodology). Projections of the adjusted pathology dataset estimate between 18,874-20,370 cases in Wales by 2030.

Costing

Based on previous works by Vallejo et al, estimated cost for diagnosis and treatment of cases identified histologically through SAIL for 2018 alone, would amount to between £6.5-£9.0 million (single case per histological type per patient). Incorporating the 45% difference calculated between pathology and SAIL data for matched regions (which represents multiple lesions data)

this figure rises to £9.5-£13.1 million. For the period 2003-2018 the total estimated cost to NHS Wales based on projected pathology data is £92-127 million. Based on the predictive model in Fig.8, the case burden in 2030 alone would equate to £18.1-24.9 million to NHS Wales as a conservative estimate (does not account for inflation and lesions treated through non-surgical methods).

Discussion

There is a paucity of robust epidemiological data in the international scientific literature related to non-melanoma skin cancer (NMSC). European guidelines[23] mandate that only a single case per histological type per patient is recorded within cancer registries – despite a 44% subsequent risk of further BCC[24]. This method of recording underestimates case burden by up to 45%, as evidenced by the discrepancy demonstrated between validated pathological data (accounting for all BCCs) and SAIL data (representing cancer registry data i.e. single lesion per patient) matched to the same region and year. This underrepresentation has clear implications for planning and resource allocation for all health services involved in the skin cancer treatment pathway. The rising global incidence of NMSC[1], coupled with struggling healthcare systems in the post COVID-19 recovery period only serve to intensify the likely societal and healthcare impact. Studies have predicted that the likely case burden for surgically managed BCCs alone in the UK will rise to 298,305 by 2025 amounting to an estimated £365 million[11,21].

Epidemiological data within Wales has been poorly represented with no previous large studies linking primary and secondary healthcare data. Recent publications proposing novel methods of data presentation have also omitted Welsh data on the basis of incomplete coverage[7]. This study represents the largest internally validated Welsh cohort examining the epidemiology of BCCs using advanced data linkage techniques to extract data from multiple routine healthcare datasets.

Based on SAIL data from the 2018 cohort, the crude incidence of BCC in Wales was 234.3/100,000 (EASR 224.6, WASR 111.8/100,000) with 7353 first recorded histologically confirmed diagnosis costing an estimated £6.5-£9.0 million.

Musah et al using the health improvement network database (THIN), a primary care repository published crude rates of 196.4/100,000 (EASR 114.4, WASR of 78.1/100000, n=2822) between 2004-2010[25]. A regional study performed by Holme and Roberts et al[5,10] highlighted a crude incidence of 224.3/100,000 (WASR 114.2/100000, n=414) in 1998, they noted a 66% increase in BCC incidence compared to a study 10 years prior. Based on the validated pathological data an increase in case burden of 105% in a 16-year period between 2003-2019 was demonstrated (average annual growth of 6.5%). Our study demonstrates one of the highest published incidences of BCCs in Europe with a higher than average rate of growth[1]. Recent studies in the Netherlands have demonstrated comparable incidences for 2018 with crude incidences rates of approximately 284 per 100,000 when counting only the first registered BCC[26]. The incidence in Wales is partially attributed to the higher proportions of Fitzpatrick skin types 1 and 2 in Wales. In the 2011 census, 96% of Welsh residents identified themselves as white[9]. Comparably there was also less inward migration of lower skin cancer susceptible individuals in comparison to other parts of the UK. Similar demographics to Wales can be found in regions such as the Southwest and Northern Ireland, but such incidences have not been reported. This, however,

remains significantly lower than published rates in Australia; the 2002 national NMSC survey published incidences of 883.7/100,000 WASR (CI 816.4-956.6)[27], with an estimated expenditure of $1,315,140,072[28] in 2019/2020.

In line with previous studies, we demonstrate a correlation between case burden and those within the least socially deprived socioeconomic group. This is partially attributable to a greater disposable income permitting international travel and outdoor recreational activities (hence greater exposure to intense intermittent UVR exposure)[29–31]. This may also be confounded by a recognised higher referral rate within the least socially deprived to specialist services[32].

Blaenau Gwent, Merthyr Tydfil, Rhondda Cynon Taf and Swansea contain the highest proportions of deprived LSOAs, however, the second most populous principle region in Wales, Swansea, displayed the highest regional incidences between 2016-2018 and Blaenau Gwent one of the lowest, suggesting that regional variations in BCC incidence cannot be explained by socioeconomic status alone[33].

A behavioural study commissioned by Welsh government published in 2009 highlighted that 8.2% of 11–17-year-olds had used a sunbed as had 22% of all respondents to the Welsh omnibus survey in 2017[34,35]. Sunbed use is significantly associated with increased risk of BCC especially under the age of 25[36].

Wales presents a unique challenge with a relatively small population situated over a large geographical region, with one in three living rurally. The geodesic distance (shortest path) to access specialist services was calculated to give an insight into distances travelled for treatment. Access to dermatological, oncological, and plastic surgery services (single journey and single lesion per patient) were considered. Over 90% of the patients were found to live within a 25mile radius of dermatological and oncological services, however this was only the case for 57% of patients seeking plastic surgery treatment, with a further 37% travelling between 25-50 miles and 6% greater than 50 miles. A systematic review of distance to healthcare services and associated implications concluded that 77% of sampled studies (n=108) displayed a distance decay association, i.e. patients living closer to healthcare facilities displayed better outcomes[37]. Considering the demographic of patients that are commonly affected, this places pressure on patients, carers and family members through missed employment - the 'Greater Patient' concept[38].

In terms of social deprivation versus access to plastic surgery tertiary care the study revealed that the least socially deprived travelled the greatest distances. This is partially attributed to the locality of the plastic surgery centre within a deprived principal region. Utilisation of healthcare services is however a complex and multifaceted issue which is not addressed by distance travelled alone.

Limitations

Limitations of this study include specific identification of BCCs from the generic ICD-10 code of C44 encompassing all NMSC. Only lesions which were histologically proven as BCC were considered using morphology coding within the WICISU dataset. The disadvantage of this, however, is that it only considers biopsy proven lesions or those surgically excised. Hence, this is likely to represent an underestimation of true burden as lesions managed through topical and destructive methods were not considered. This is confounded by the recognised underestimation of using the first diagnosed BCC per patient. This effect is carried forward when

linked to OPDW and PEDW databases, however without linkage techniques, coding would be limited to generic ICD-10 coding and therefore specific BCC delineation would not be possible from the umbrella of NMSC. This was partially circumvented by integration of pathological data scaled to the Welsh population as this considered all histologically confirmed BCC not just the first registered. Due to limited historic records, reporting the first diagnosed BCC per patient may falsely represent their first lesion within a year that does not actually represent the first diagnosis of BCC in their lifetime. Other limitations include missing, incomplete, and misclassified datasets. GIS calculations were based on Geodesic distances from LSOA to treatment centre. The modifiable area unit problem (MAUP), a form a statistical bias when using spatial analysis must be recognised, this occurs when varying configurations of spatial units cause modified aggregation of individuals[39,40]. Standardised populations within WASR are less comparable to the population of Wales, as they are based on a younger population structure, hence the noted disparity between the EASR and WASR.

Conclusion

This study is the first to present the burden of BCC in Wales using the largest population cohort to date, supported by regional pathological data for validation and forecasting. This study demonstrates one of the highest age standardised incidences in Europe and forecasts a conservative growth in case burden of approximately 25% in the next 10 years and a three-fold increase in the expected expenditure between 2018 and 2030. Although there have been published improvements in recent years in registration technique, we cannot rely on historical coding to predict future service needs as this will result in chronic under resourcing. If left unchecked the burden of BCC within the umbrella of NMSC will likely overwhelm future services, in addition to the strain placed in the post Covid-19 recovery, resulting in a significant impact on waiting lists for both assessment and treatment of this disease.

Acknowledgements

The authors would like to thank Mr Rhys Whelan, Library Services Manager, Swansea Bay University Health Board, Morriston Hospital, Swansea, SA6 6NL, UK for his support with the literature search. The Authors would also like to thank Mr Octavian Parkes for his support as Research Programme Manager.

Funding: This research was supported via the RESECT project, which is part of the Scar Free Foundation Programme of Regenerative Research at the Reconstructive Surgery & Regenerative Medicine Research Centre (ReconRegen) in partnership with Health and Care Research Wales. This work was supported by Health Data Research UK, which receives its funding from HDR UK Ltd (HDR-9006) funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation (BHF) and the Wellcome Trust; ADR Wales is

part of the Economic and Social Research Council (part of UK Research and Innovation) funded ADR UK (grant ES/S007393/1).

Conflicts of interest: None to declare

Data availability: The data was acquired from the SAIL Databank at Swansea University, Swansea, UK, but as restrictions apply, they are not publicly available.

Ethics statement: Approval for the use of anonymised data in this study, provisioned within the SAIL Databank was granted by an independent Information Governance Review Panel (IGRP) under project 0593. The IGRP has a membership comprised of senior representatives from the British Medical Association (BMA), the National Research Ethics Service (NRES), Public Health Wales and NHS Wales Informatics Service (NWIS). Usage of additional data was granted by data owner. The SAIL Databank is General Data Protection Regulations (GDPR) and the UK Data Protection Act compliant.

What is already known about this topic?

- The incidence of basal cell carcinoma is recognised to be increasing
- Epidemiological studies in Wales are limited with no significant large cohort basal cell carcinoma research performed

What does this study add?

- This study identifies the incidence of basal cell carcinoma in Wales and recognises it as one of the highest published in Europe.
- This study identifies geographical and social variations, whilst integrating routinely collected healthcare data to assess distances travelled for specialist input.
- This study provides an assessment of the significant financial implication of care in Wales, and a projection of future burden of care to inform service planning.

References

1 Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. *Br J Dermatol* 2012; **166**:1069–80.

2 Ibrahim N, Gibson J, Ali S, *et al.* Is poor quality non-melanoma skin cancer data affecting high quality research and patient care? *J Plast Reconstr Aesthetic Surg* 2020. doi:10.1016/j.bjps.2020.12.036.

3 Lucke TW, Hole DJ, Mackie RM. An audit of the completeness of non-melanoma skin cancer registration in Greater Glasgow. *Br J Dermatol* 1997; **137**:761–3.

4 BEADLE PC, BULLOCK D, BEDFORD G, *et al.* Accuracy of skin cancer incidence data in the United Kingdom. *Clin Exp Dermatol* 1982; **7**:255–60.

5 Roberts DL. Incidence of non-melanoma skin cancer in West Glamorgan, South Wales. *Br J Dermatol* 1990; **122**:399–403.

6 de Vries E. Population-Based Estimates of the Occurrence of Multiple vs First Primary Basal Cell Carcinomas in 4 European Regions. *Arch Dermatol* 2012; **148**:347.

7 Venables ZC, Nijsten T, Wong KF, *et al.* Epidemiology of basal and cutaneous squamous cell carcinoma in the U.K. 2013–15: a cohort study. *Br J Dermatol* 2019; **181**:474–82.

8 Brewster DH, Bhatti LA, Inglis JHC, *et al.* Recent trends in incidence of nonmelanoma skin

cancers in the East of Scotland, 1992–2003. *Br J Dermatol* 2007; **156**:1295–300.
9 Consensus [WWW Document]. 2011.URL https://www.nomisweb.co.uk/census/2011/KS201EW/view/2092957700?cols=measures [accessed on 5 December 2020].
10 Holme SA, Malinovszky K, Roberts DL. Changing trends in non-melanoma skin cancer in South Wales, 1988–98. *Br J Dermatol* 2000; **143**:1224–9.
11 Goon PKC, Greenberg DC, Igali L, Levell NJ. Predicted cases of U.K. skin squamous cell carcinoma and basal cell carcinoma in 2020 and 2025: horizon planning for National Health Service dermatology and dermatopathology. *Br J Dermatol* 2017; **176**:1351–3.
12 Benchimol EI, Smeeth L, Guttmann A, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Med* 2015; **12**:e1001885.
13 Ford D V, Jones KH, Verplancke J-P, *et al.* The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009; **9**:157.
14 Jones KH, Ford D V., Jones C, *et al.* A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform* 2014; **50**:196–204.
15 Lyons RA, Jones KH, John G, *et al.* The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009; **9**:3.
16 Government W. StatsWales-Welsh Index of Multiple Deprivation [WWW Document]. URL https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation [accessed on 6 April 2021].
17 Eurostat. Revision of European Standard Population [WWW Document]. 2013.
18 Waterhouse, J., Muir, C.S., Correa, P., Powell J. Cancer Incidence in Five Continents. *IARC Sci Publ No 15, Lyon, IARC* 1976; **Vol. III**.
19 Ahmad OB, Boschi-Pinto C, Lopez AD, *et al.* AGE STANDARDIZATION OF RATES: A NEW WHO STANDARD. *World Heal Organ* 2001; **GPE Discus**.URL https://www.who.int/healthinfo/paper31.pdf.
20 Welsh Government. Stats Wales [WWW Document]. .
21 Vallejo-Torres L, Morris S, Kinge JM, *et al.* Measuring current and future cost of skin cancer in England. *J Public Health (Bangkok)* 2014; **36**:140–8.
22 Government WA. National Standards for Skin Cancer Services. , 2005.
23 ENCR. ENCR Recommendations-Non melanoma skin cancer [WWW Document]. 2000.URL https://www.encr.eu/sites/default/files/pdf/skinrecs.pdf [accessed on 5 December 2020].
24 Marcil I, Stern RS. Risk of Developing a Subsequent Nonmelanoma Skin Cancer in Patients With a History of Nonmelanoma Skin Cancer. *Arch Dermatol* 2000; **136**. doi:10.1001/archderm.136.12.1524.
25 Musah A, Gibson JE, Leonardi-Bee J, *et al.* Regional variations of basal cell carcinoma incidence in the U.K. using The Health Improvement Network database (2004-10). *Br J Dermatol* 2013; **169**:1093–9.
26 Schreuder K, Hollestein L, Nijsten TEC, *et al.* A nationwide study of the incidence and trends of first and multiple basal cell carcinomas in the Netherlands and prediction of future incidence. *Br J Dermatol* 2021. doi:10.1111/bjd.20871.
27 Initiative NCC, Group NSCW. The 2002 national non-melanoma skin cancer survey. ,

2002URL https://www.canceraustralia.gov.au/publications-and-resources/cancer-australia-publications/2002-national-non-melanoma-skin-cancer-survey.
28 Australian Institute of Health and Welfare. Health Expenditure Australia 2019-20 [WWW Document]. 2021.URL https://www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure-australia-2019-20/contents/about [accessed on 31 January 2022].
29 Gallagher RP, Hill GB, Bajdik CD, *et al.* Sunlight Exposure, Pigmentary Factors, and Risk of Nonmelanocytic Skin Cancer: I. Basal Cell Carcinoma. *Arch Dermatol* 1995; **131**:157–63.
30 Kricker A, Armstrong BK, English DR, Heenan PJ. Does intermittent sun exposure cause basal cell carcinoma? a case-control study in Western Australia. *Int J Cancer* 1995; **60**:489–94.
31 Rosso S, Zanetti R, Martinez C, *et al.* The multicentre south European study 'Helios'. II: Different sun exposure patterns in the aetiology of basal cell and squamous cell carcinomas of the skin. *Br J Cancer* 1996; **73**:1447–54.
32 van Doorslaer E. Inequalities in access to medical care by income in developed countries. *Can Med Assoc J* 2006; **174**:177–83.
33 Welsh Government. Welsh Index of Multiple Deprivation (WIMD) 2019: Results report. , 2019.
34 Government W. The sunbed regulation act 2011. 2011; **2010**.URL https://gov.wales/sites/default/files/publications/2019-04/the-sunbeds-regulation-act-2010-wales-regulations-2011.pdf.
35 Knowledge and Analytical services WG. Survey of sunbed use in Wales : Summary briefing. 2017.
36 Wehner MR, Shive ML, Chren M-M, *et al.* Indoor tanning and non-melanoma skin cancer: systematic review and meta-analysis. *BMJ* 2012; **345**:e5909–e5909.
37 Kelly C, Hulme C, Farragher T, Clarke G. Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? A systematic review. *BMJ Open* 2016; **6**:e013059.
38 Basra MKA, Finlay AY. The family impact of skin diseases: the Greater Patient concept. *Br J Dermatol* 2007; **156**:929–37.
39 Openshaw S. Ecological Fallacies and the Analysis of Areal Census Data. *Environ Plan A Econ Sp* 1984; **16**:17–31.
40 Gehlke CE, Biehl K. Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. *J Am Stat Assoc* 1934; **29**:169–70.

Figure legends

Figure 1 Data preparation of the e-cohort

Figure 2 European age standardised incidence (per 100,000) of BCC Male/Female 2000-2018. Shading denoting upper and lower 95% confidence intervals.

Figure 3 Age specific incidence (per 100,000 PY) per decade age group 2016-2018.

Figure 4 Depicting original and adjusted SAIL crude incidence for Wales 2003-2018.Original SAIL data depicted in blue, the blue shading outlines the period of non-mandatory data collection within WICSU followed by an increase in 2016-2018 when mandatory collection was introduced. The adjusted crude incidence (solid orange line) has been calculated by determining the relationship between regional pathological data and Pan Wales SAIL data 2016-2018. This has been backward projected by multiplying this relationship by the adjusted pathological data to give the adjusted SAIL data, i.e. a more accurate reflection of pan Wales incidence during the non-mandatory skin cancer registration period (NB regional data is based on histopathological reports and hence was not effected by cancer registry practice). The orange dashed line represents the line of best fit through the forecasted SAIL data.

Figure 5 BCC caseload based on Urban/ Rural locality 2016-2018

Figure 6 BCC Crude incidence per 100,000 (PY) based on Urban/Rural locality (Morphology codes 1-3) 2016-2018

Figure 7 EASR of BCC in each principal region within Wales 2016-2018 (per 100,000 PY)

Figure 8 Projection of adjusted SBUHB cases (Swansea and Neath Port Talbot) to the population of Wales to 2030 (Blue). 3 potential projections, based on utilisation of Linear (Green dashed line), Quadratic (Red dashed line) and a “expected” scenario derived by taking an average of both models (Blue dashed line). Projection of adjusted SAIL data to 2030 (Purple). 3 potential projections demonstrated, Linear (Light blue dashed line), Quadratic (orange dashed line) and “expected”, an average of both models (purple dashed line). Multiple projection models in SPSS were tested, we concluded the best two models were that of a linear and non – linear (quadratic) equation. The linear model with a constant gradient was used to model the best-case scenario, i.e. that the cases do not change by a significant amount year on year. The quadratic model with the x^2 term clearly has a non-uniform gradient, the predicted cases can differ significantly year on year, and thus was used it to model the worst-case scenario. From this an expected model was formulated defined as the average of the worst-case and best-case models.

Database Name	Description
Patient Episode Database for Wales (PEDW)	Attendance and clinical information for all hospital admission including operations performed in Wales.
Outpatient Dataset for Wales (OPDW)	Attendance information for all hospital outpatient appointments in Wales.
Welsh Cancer Intelligence and Surveillance Unit (WCISU)	Cancer Registry for Wales, used automated systems to integrate information from multiple sources to include Histopathology, MDT data etc.
Welsh Longitudinal General Practice Dataset (WLGP)	Attendance and clinical information for all general practice interactions.
Welsh Demographic Service Dataset (WDSD)	Population spine for Wales, used to identify anonymised changes in general practice registration and residency over time.

Table 1 Definitions of accessed data sources

Variable	Males n=33,440	Females n=27,909
Age Group n (%)	(54.5)	(45.5)
5-19*	15	12
20-24	33(0.1)	29(0.1)
25-29	64(0.2)	82(0.3)
30-34	148(0.4)	184(0.7)
35-39	275(0.8)	380(1.4)
40-44	539(1.6)	643(2.3)
45-49	903(2.7)	1064(3.8)
50-54	1542(4.6)	1555(5.6)
55-59	2288(6.8)	1984(7.1)
60-64	3309(9.9)	2540(9.1)
65-69	4766(14.3)	3456(12.4)
70-74	5544(16.6)	3867(13.9)
75-79	5742(17.2)	4014(14.4)
80-84	4673(14.0)	3842(13.8)
85-90	2874(8.6)	3100(11.1)
>90	725(2.2)	1157(4.1)
Grand Total	33440	27909
WIMD 2011		
1 (Most Deprived)	4732(14.2)	4155(14.9)
2	5710(17.1)	4891(17.5)
3	6608(19.8)	5571(20)
4	7464(22.3)	6229(22.3)
5 (Least Deprived)	8846(26.5)	6990(25.1)
NA	80(0.2)	73(0.3)
Site		
Lip	342(1.0)	650(2.3)
Eyelid/Canthus	2432(7.3)	2617(9.4)
Ear	2687(8.0)	449(1.6)
Face	11683(34.9)	10884(39)
Scalp/Neck	1808(5.4)	1366(4.9)
Trunk	2811(8.4)	1508(5.4)
Upper Limb	1228(3.7)	782(2.8)
Lower Limb	729(2.2)	1829(6.6)
Overlapping lesion	40(0.1)	32(0.1)
Unspecified	9680(28.9)	7792(27.9)

*Table 2 Patient Demographics 2000-2018 *Age groups merged to avoid risk of unmasking when n<5, as per SAIL disclosure control policies*

Speciality	Proportion of outpatient appointments with a diagnosis of BCC	Proportion of outpatient appointments with C44 code
Dermatology	92.9%	89.5%
Plastic Surgery	3.0%	5.8%
Clinical Oncology	3.7%	4.0%

*Table 3 Proportion of appointments per speciality using both generically coded C44 data and data linkage from OPDW to WCISU 2004-2018 *NB remaining are associated with allied surgical specialities*

Speciality	Proportion of inpatient/Day case spell
Dermatology	40.5%
Plastic Surgery	23.4%
Oral Surgery	11.8%
Clinical Oncology	8.0%
Ophthalmology	7.7%
Ear, Nose and Throat	5.0%
General Surgery	3.0%

*Table 4 Proportion of Inpatient/Day case episodes by speciality 2000-2018-PEDW Database *NB remaining are associated with allied surgical specialities*

Distance Travelled	Dermatology	Plastic Surgery	Oncology
<25 Miles	97.9%	57.1%	91.6%
≥25-50 Miles	1.8%	36.9%	7.4%
≥50 Miles	0.3%	6%	1%

Table 5 Access to dermatology, plastic surgery and oncological services 2000-2018. Geographical reference points: Wales from north to south spans 130 miles and the widest point spans 100 miles from east to west.

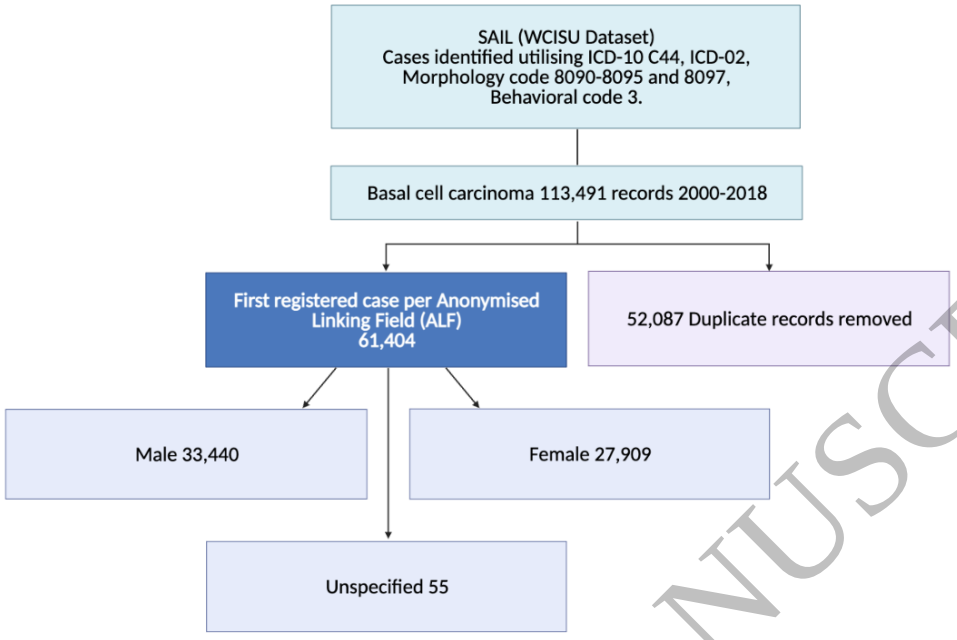

Figure 1 Data preparation of the e-cohort

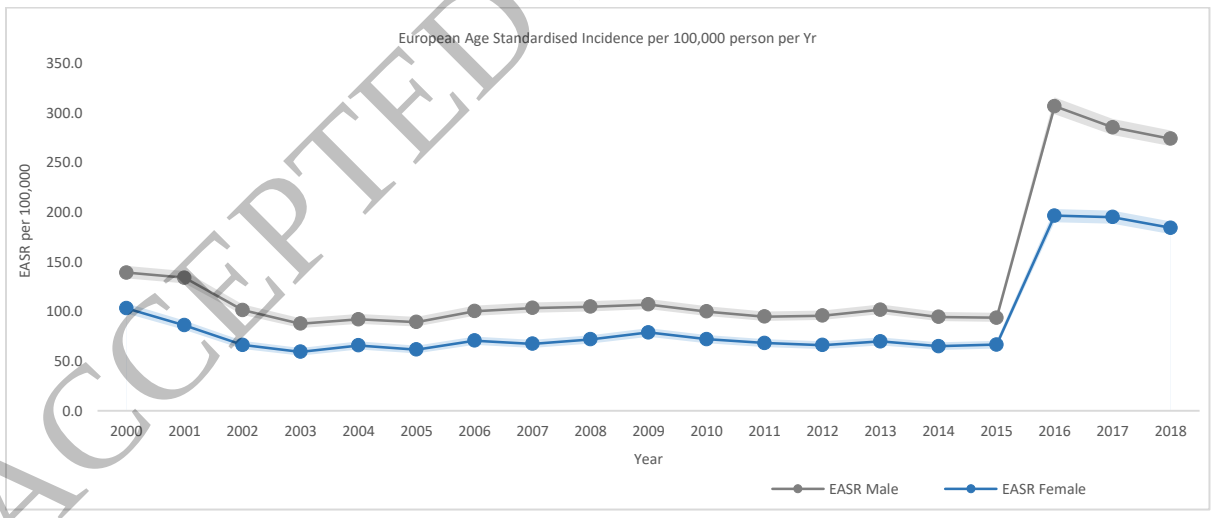

Figure 2 European age standardised incidence (per 100,000) of BCC Male/Female 2000-2018. Shading denoting upper and lower 95% confidence intervals.

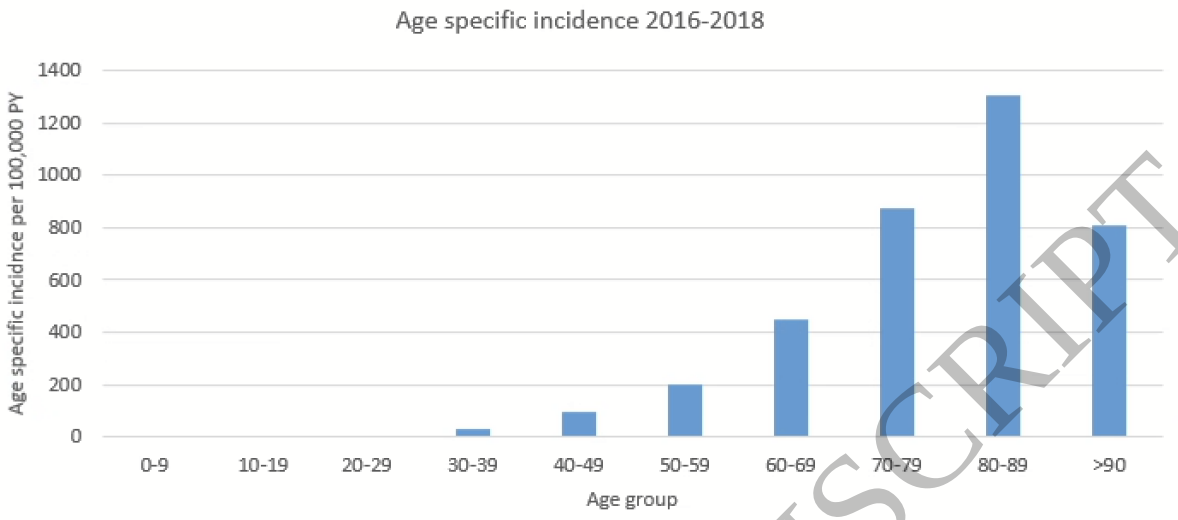

Figure 3 Age specific incidence (per 100,000 PY) per decade age group 2016-2018.

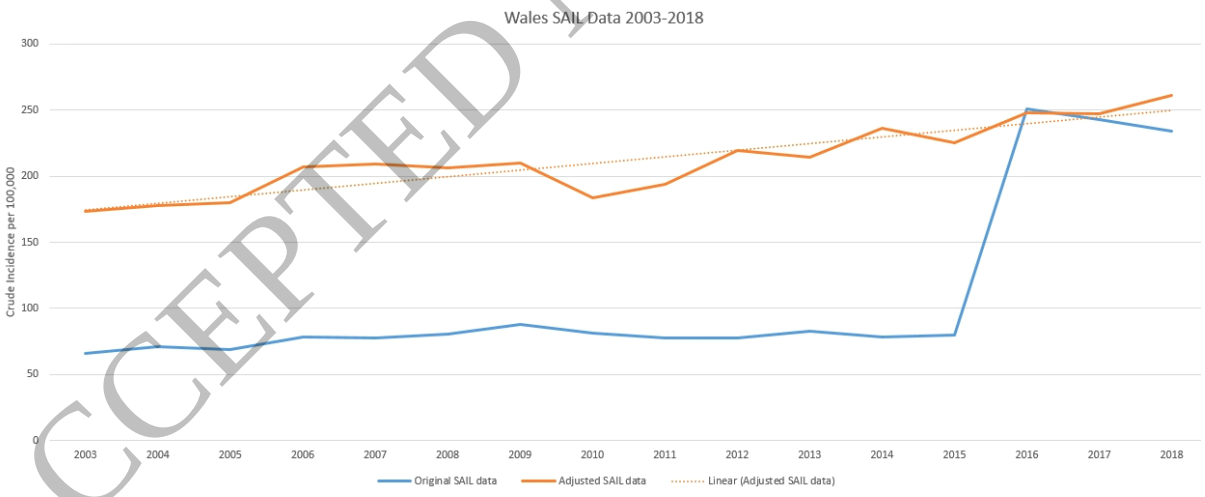

Figure 4 Depicting original and adjusted SAIL crude incidence for Wales 2003-2018.Original SAIL data depicted in blue, the blue shading outlines the period of non-mandatory data collection within WICSU followed by an increase in 2016-2018 when mandatory collection was introduced. The adjusted crude incidence (solid orange line) has been calculated by determining the relationship between regional pathological data and Pan Wales SAIL data 2016-2018. This has been backward projected by multiplying this relationship by the adjusted pathological data to give the adjusted SAIL data, i.e. a more accurate reflection of pan Wales incidence during the non-mandatory skin cancer registration period (NB regional data is based on histopathological reports and hence was not effected by cancer registry practice). The orange dashed line represents the line of best fit through the forecasted SAIL data.

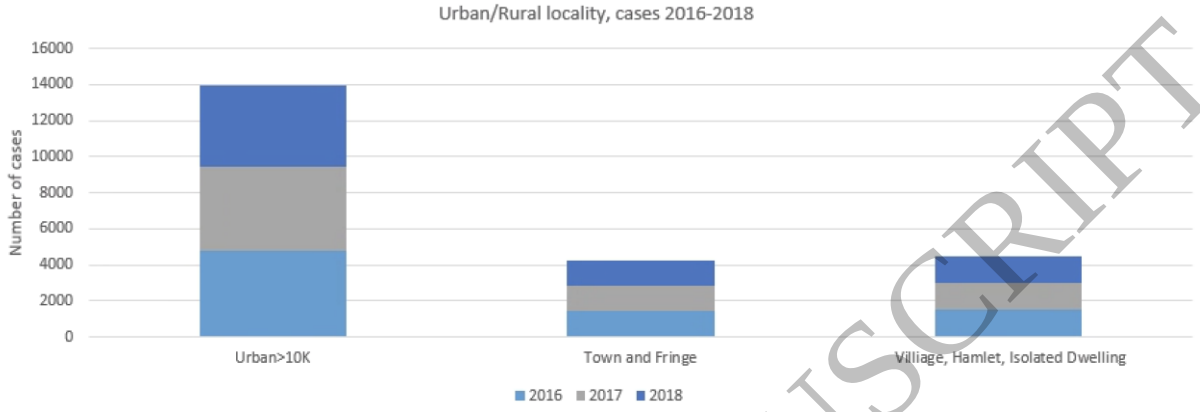

Figure 5 BCC caseload based on Urban/ Rural locality 2016-2018

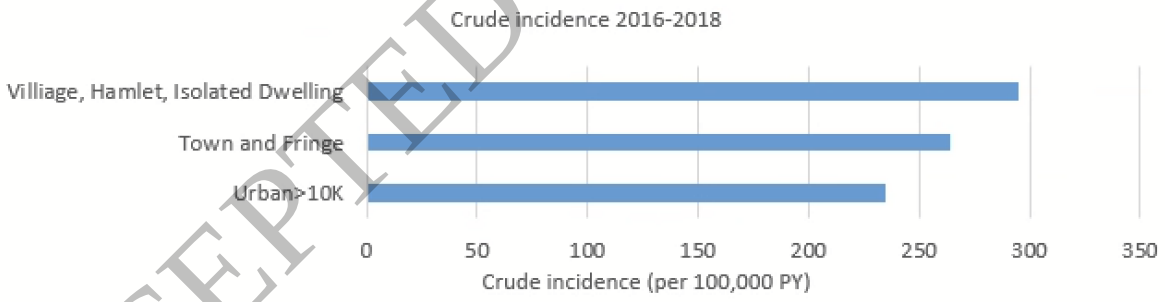

Figure 6 BCC Crude incidence per 100,000 (PY) based on Urban/Rural locality (Morphology codes 1-3) 2016-2018

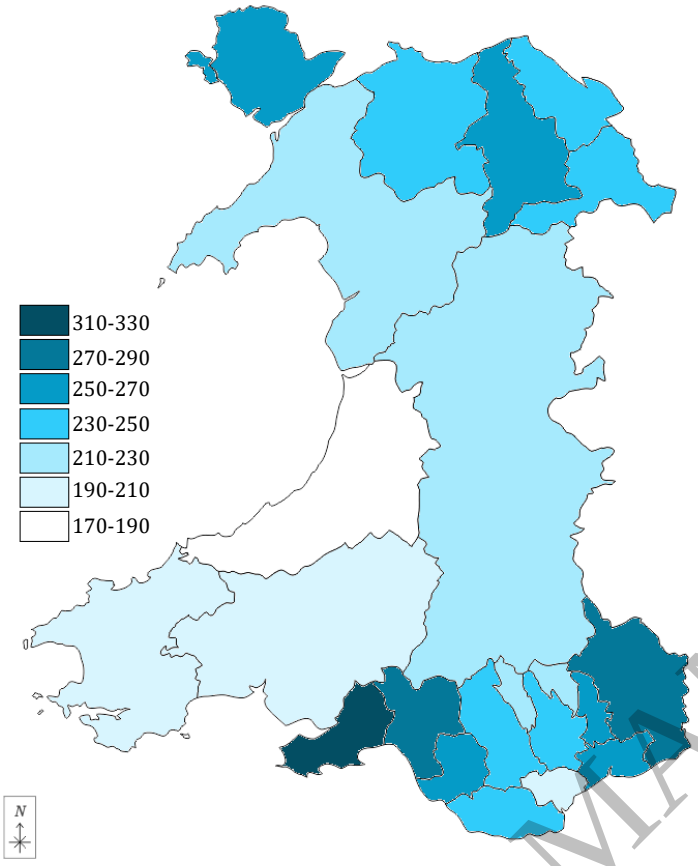

Figure 7 EASR of BCC in each principal region within Wales 2016-2018 (per 100,000 PY)

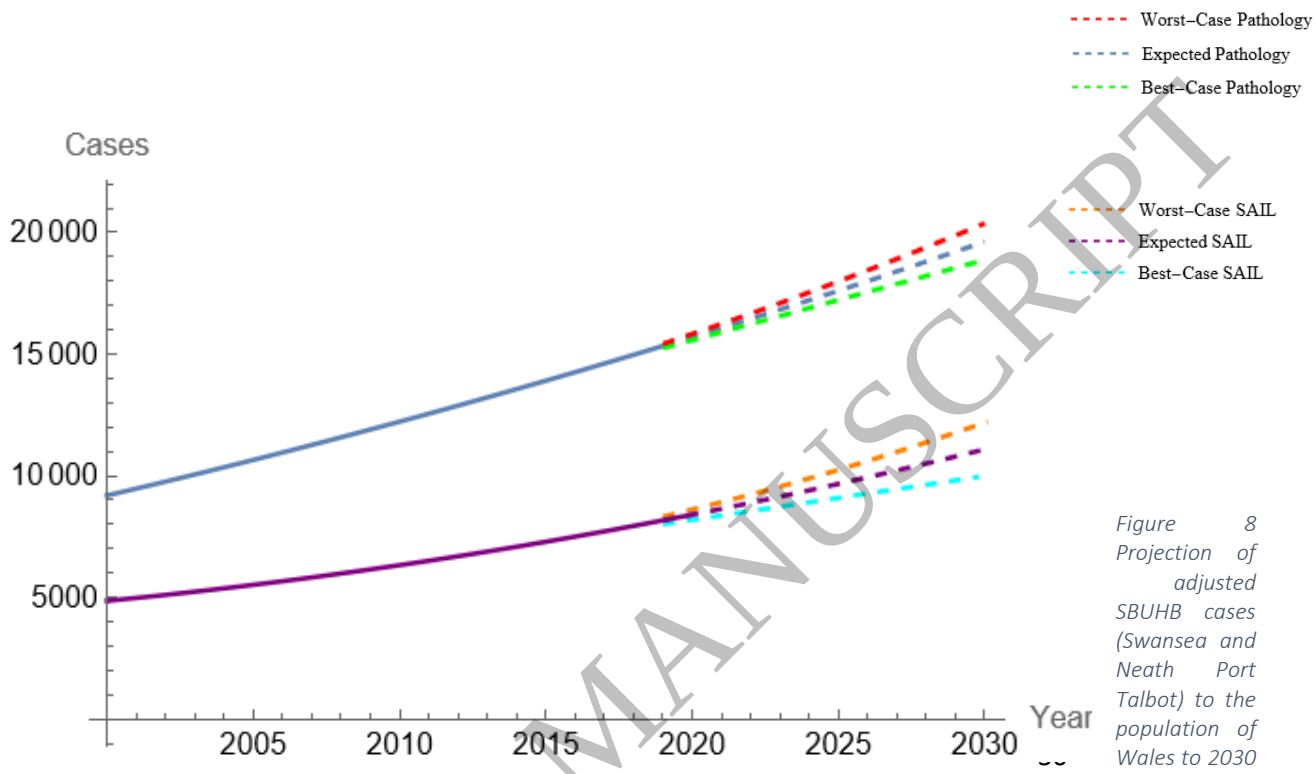

Figure 8 Projection of adjusted SBUHB cases (Swansea and Neath Port Talbot) to the population of Wales to 2030 (Blue). 3 potential projections, based on utilisation of Linear (Green dashed line), Quadratic (Red dashed line) and a "expected" scenario derived by taking an average of both models (Blue dashed line). Projection of adjusted SAIL data to 2030 (Purple). 3 potential projections demonstrated, Linear (Light blue dashed line), Quadratic (orange dashed line) and "expected", an average of both models (purple dashed line). Multiple projection models in SPSS were tested, we concluded the best two models were that of a linear and non – linear (quadratic) equation. The linear model with a constant gradient was used to model the best-case scenario, i.e. that the cases do not change by a significant amount year on year. The quadratic model with the x^2 term clearly has a non-uniform gradient, the predicted cases can differ significantly year on year, and thus was used it to model the worst-case scenario. From this an expected model was formulated defined as the average of the worst-case and best-case models.