



Cancer Burden in Belgium: the good, the bad, (the grey) and the ugly

Data collection,
data validation and data cleaning

Frédéric Calay
10/12/2015

10 years

Data collection - Data processing



Data collection

Formatting

Conversions

Import

Feedback

Algorithm

Quality control by data providers

Reporting

Data validation: routine process

**Data cleaning for
specific purposes**

Data collection



104 Hospitals
100,000 records per year

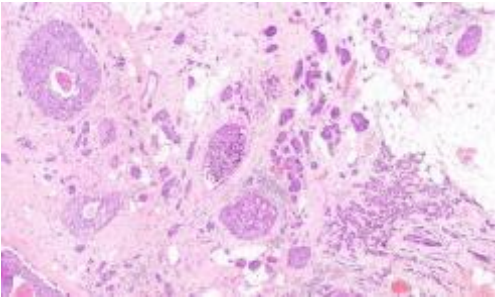
Web Based Cancer
Registration (WBCR)

sFTP

sFTP

sFTP: Secure File Transfer Protocol

**92 Laboratories for
pathological anatomy**
650,000 records per year



Registration
form

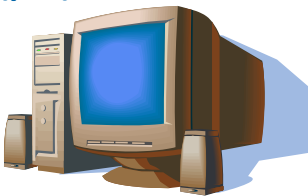


Health insurance companies

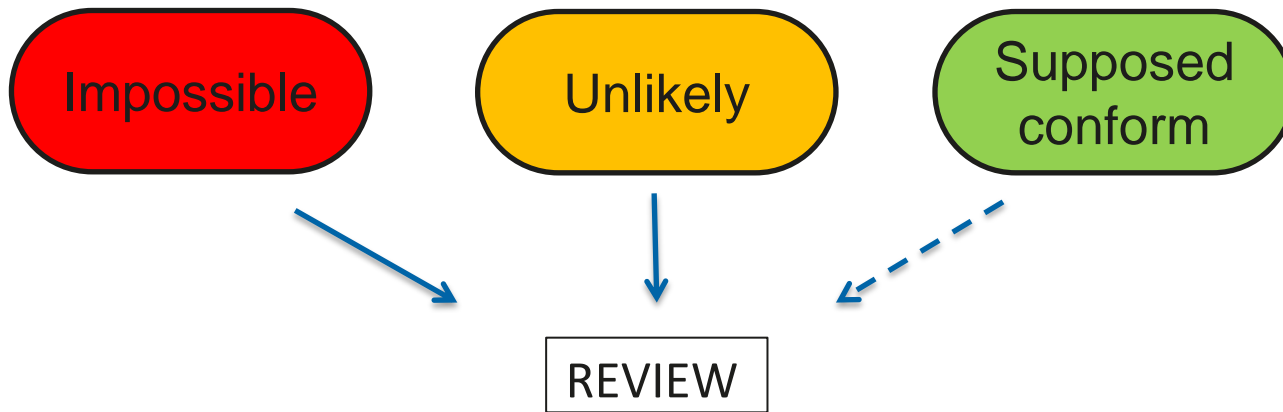
Registration
form

**750,000 records
per year!**

Historic database
1,350,000 (pre) invasive tumours



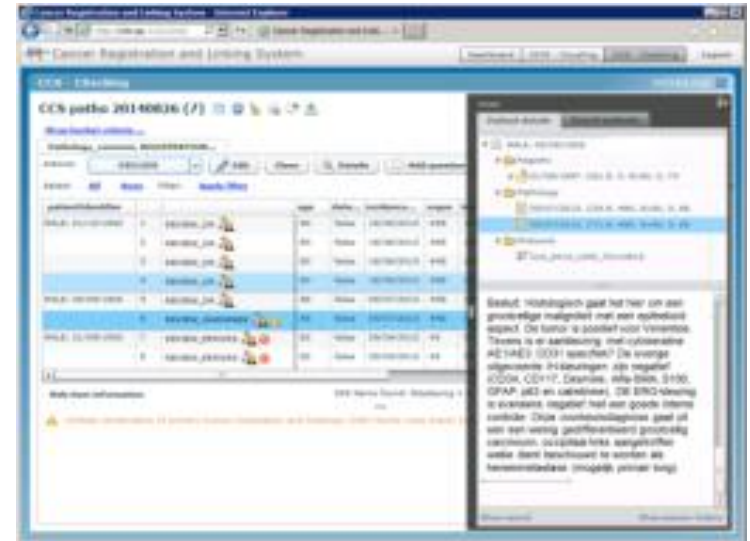
Data validation



Decision making requires expertise

Per year: manual review of:

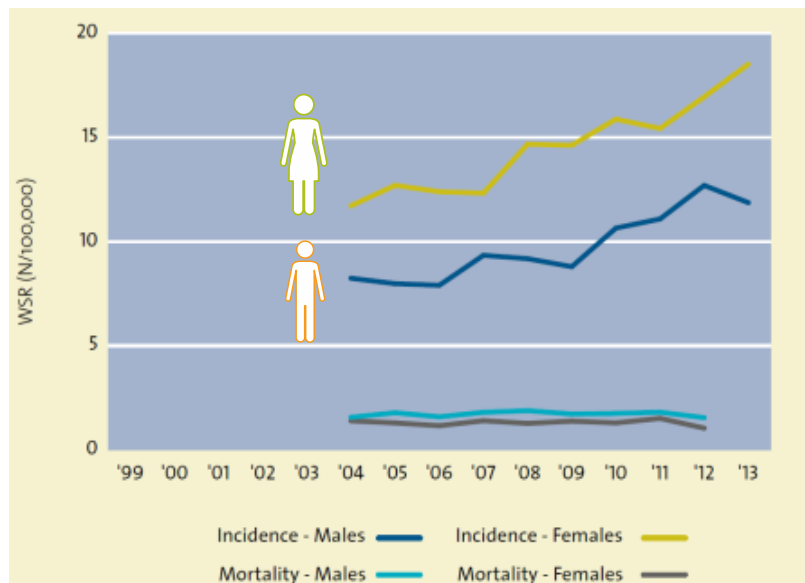
- ~ **10.500 errors**
- ~ **12.000 warnings**
- ~ **43.000 valid cases**



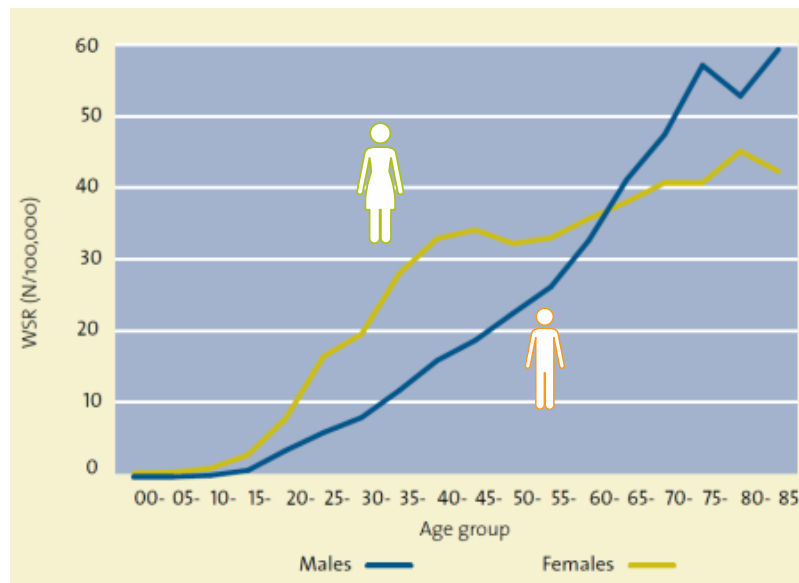
Specific data cleaning: Malignant melanoma



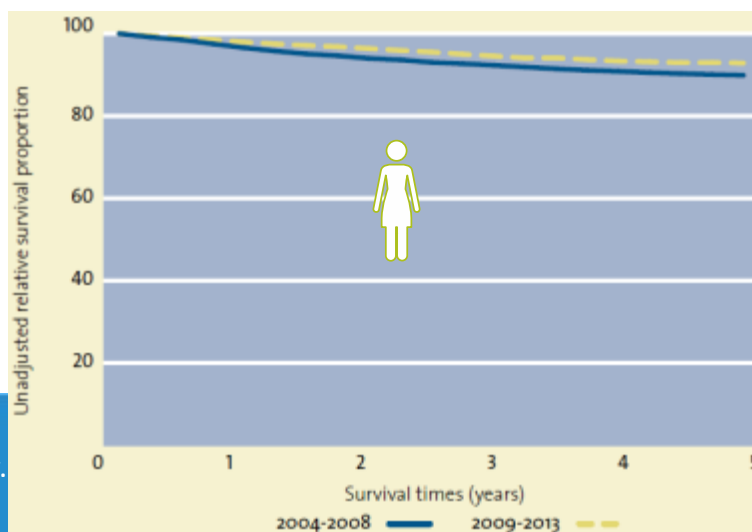
Trends in age-standardised incidence and mortality (WSR) by sex, 2004-2013



Age-specific incidence rates (N/100,000) by sex, 2009-2013



Relative survival for females by cohort in Belgium 2004-2008, 2009-2013

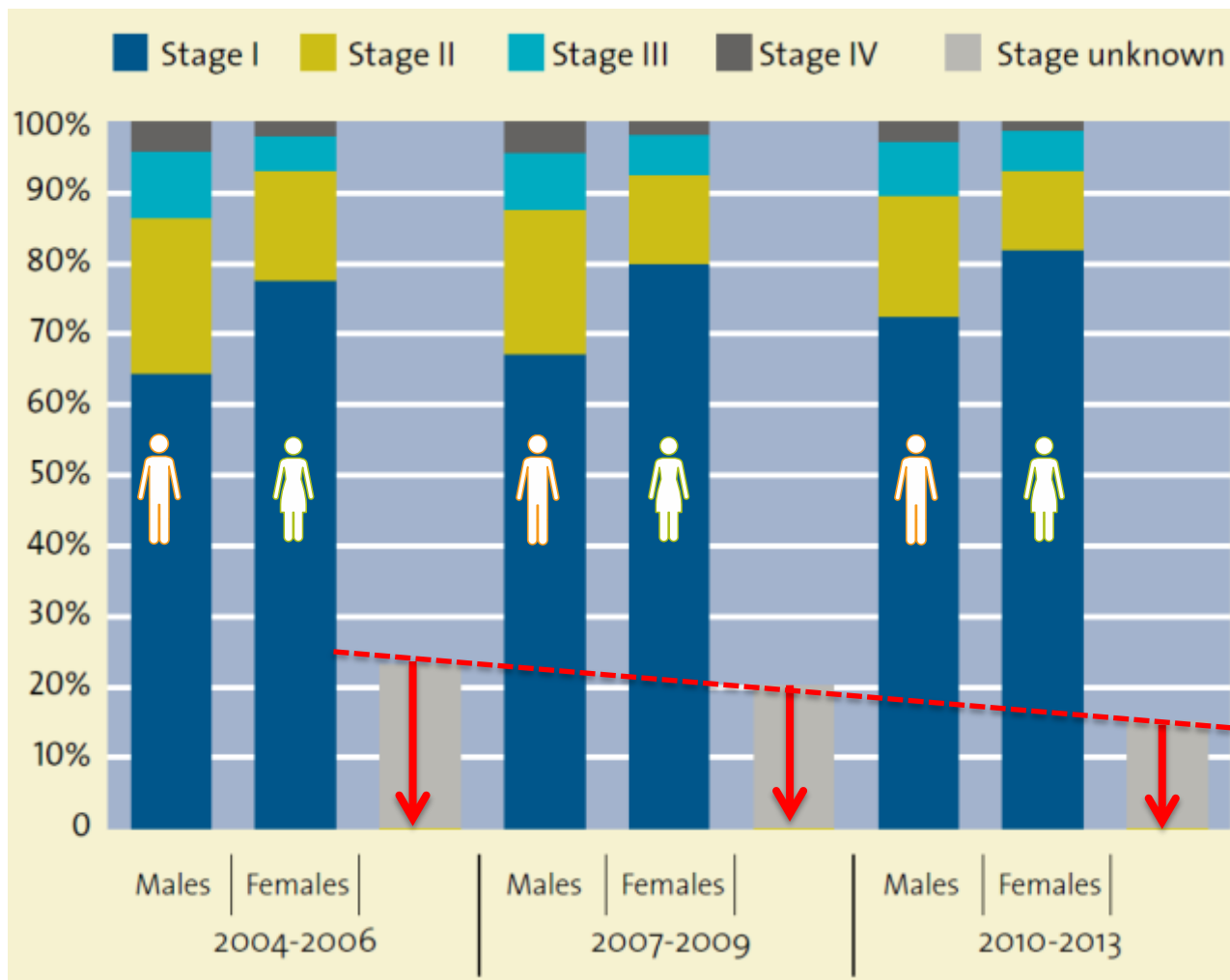


2009-2013
2004-2008

Specific data cleaning: Malignant melanoma



Stage distribution by sex, 2004-2006, 2007-2009 and 2010-2013

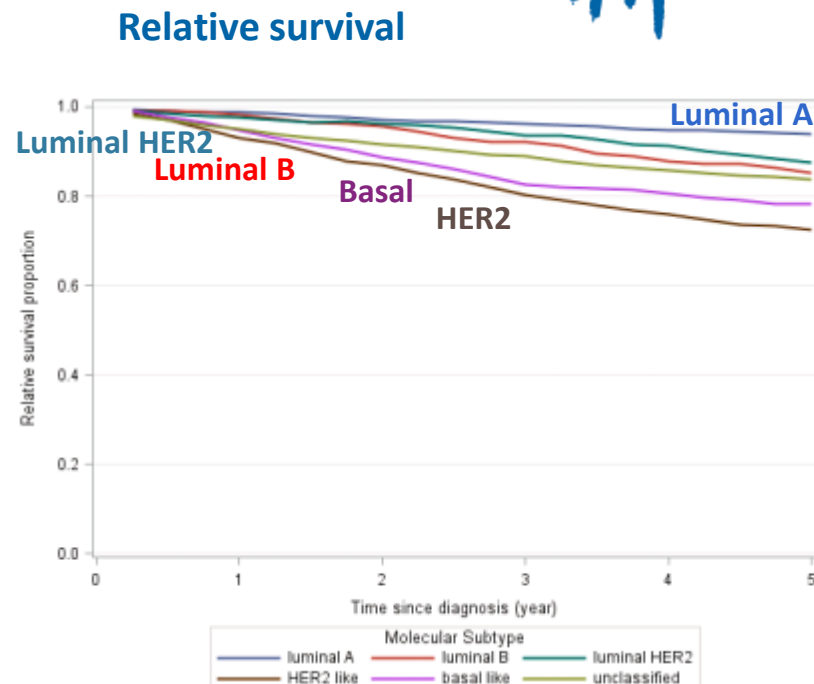


Specific data cleaning: Breast cancer



Retrospective study on molecular markers in pathological reports 2008 (N=10,007)

Breast cancer subtype	characteristics	%
Luminal A-like	ER+ or PR+/ grade 1 or 2/ HER2-	44.3 %
Luminal B-like	ER+ or PR+/grade 3/HER2-	12.8%
Luminal HER2	ER+ or PR+/HER2+	9.7%
HER2-like	ER-/PR-/HER2+	3.8%
Basal like	ER-/PR-/HER2-	8.2%
Unclassified	Too few information available	21.3%

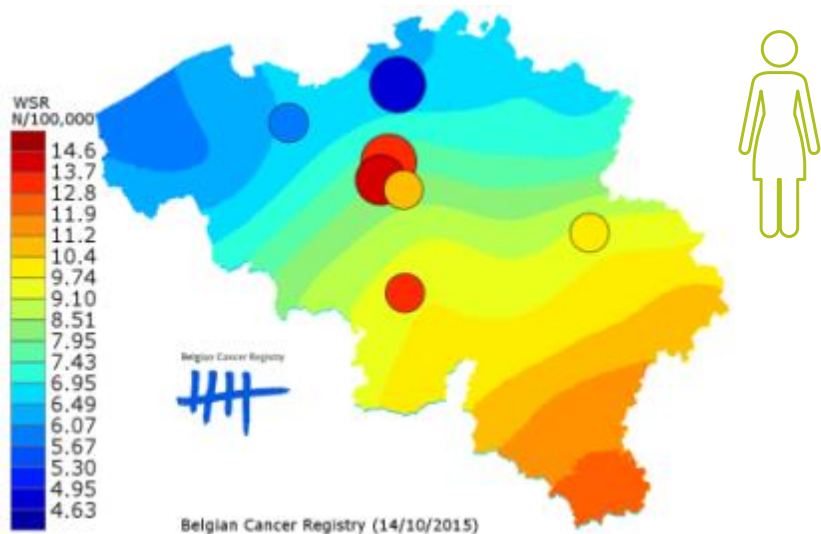


Quality of pathology reporting is crucial for cancer care and registration: A baseline assessment for breast cancers diagnosed in Belgium in 2008

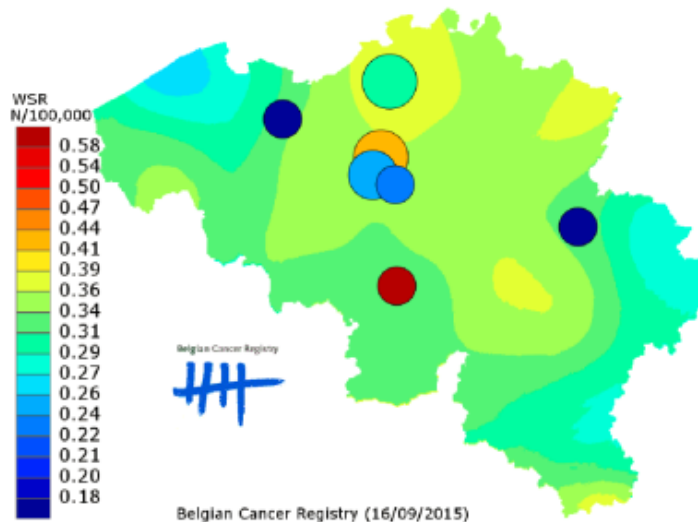
H. De Schutter ^{a, *}, N. Van Damme ^a, C. Colpaert ^b, C. Galant ^c, K. Lambein ^d, A. Cornelis ^e, P. Neven ^f, E. Van Eycken ^a

Specific data cleaning: Thyroid cancer

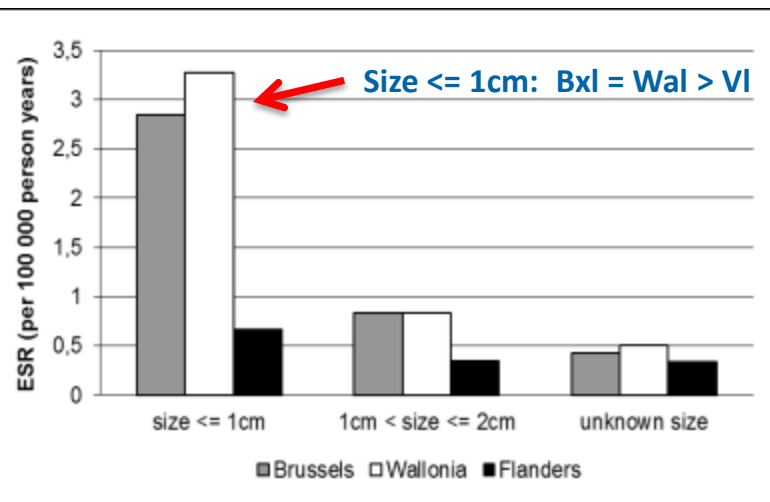
Incidence 2004-2013, Female



Mortality 2004-2012, Female



Incidence of T1 papillary carcinomas by size, 2004-2006



REGIONAL DIFFERENCES IN THYROID CANCER INCIDENCE IN BELGIUM: ROLE OF DIAGNOSTIC AND THERAPEUTIC STRATEGIES FOR THYROID DISEASE



Take home messages



- Collaboration with 104 Hospitals, 92 Laboratories and Health insurance companies
- 750,000 records per year lead to ~ 100,000 (pre) invasive tumours
- Additional information from pathological reports
→ supporting research projects