



**SMALL NUMBERS PUBLISHING GUIDANCE**

**Introduction and Aim**

This document sets out the approach to disclosure control in Public Health Wales when publishing statistical data/information. It includes a summary of approaches for handling small number data/information. It includes a flow chart for adaption by data controllers of specific data sets.

**Linked Policies, Procedures and Written Control Documents**

Information Governance Framework

**Scope**

Anyone in Public Health Wales that creates information that contain small numbers, i.e. potentially identifying the subject.

**Equality and Health Impact Assessment (EHIA)**

This guidance was written to ensure that the data of anyone with special circumstances was handled in such a way that it was not disclosable. Thereby ensuring that producers or users of data advance equality, eliminate discrimination and foster good relationships. It was determined that a more comprehensive EHIA was not required.

**Approved by**

Information Governance Working Group

**Approval Date**

31 July 2017

**Review Date**

October 2018

**Date of Publication:**

8 August 2017

**Group with authority to approve supporting procedures**

Information Governance Working Group

**Accountable Executive Director/Director**

Rhiannon Beaumont-Wood, Executive Director of Quality, Nursing and Allied Health Professionals

**Author**

Sub group of data handlers for Information Governance Working Group

**Disclaimer**

**If the review date of this document has passed please ensure that the version you are using is the most up to date either by contacting the document author or the [Corporate Governance](#).**

**Summary of reviews/amendments**

| <b>Version number</b> | <b>Date of Review</b> | <b>Date of Approval</b> | <b>Date published</b> | <b>Summary of Amendments</b>   |
|-----------------------|-----------------------|-------------------------|-----------------------|--|
| V1                    | --                    | 24.2.12                 | --                    | Sub group of the Information Governance Working Group (IGWG) developed this document as a guide to staff preparing data and staff using data. This was approved at IGWG as an internal document. |
| V2                    | 29.9.15               | 31.07.17                | 8.8.17                | Complete revamp setting out flow charts for use by Teams/Divisions preparing and using data.   |
| V2.1                  |                       |                         | 31.08.22              | Numbering updated to reflect accurate naming convention from PHW48 to AW16 CD03  |

## **1 Introduction**

This document is a guide for Public Health Wales staff who produce information in tabular form covering the principles and procedures for safeguarding the confidentiality and managing the disclosure of patient/staff information that is actually or potentially identifiable.

Disclosure of data/information which could potentially identify individuals that can cause harm, breach trust and break legal and contractual obligations. Risk of potentially identifiable data/information getting into the public domain would be a reputational risk to the organisation.

## **2 What type of data/information does this apply to**

This relates to raw data relating to individuals that could be potentially identifiable.

Data can be owned and controlled by someone other than Public Health Wales e.g. Office for National Statistics (ONS), NHS Wales Informatics Service (NWIS) or Welsh Government (WG) or the data may be controlled by Public Health Wales.

Public Health Wales has access to a number of datasets which are owned and managed by NWIS. The main dataset is the Patient Episode Database for Wales (PEDW).

Operational use of data, e.g. control of outbreaks of infectious disease – this activity is outside the remit of this guidance.

Data of individual patients can normally be shared for clinical purposes i.e. consultant led audit of unit/departmental data.

### **2.1 Disclosure control considerations**

Data/information has to be released at a high enough level of aggregation to prevent others from recognising a particular individual.

Small numbers are not necessarily a problem when they cover a broad geographical area, because the patient/staff would not normally be identifiable. However, data/information that are likely to be more sensitive, e.g. deaths, abortions or specific conditions should still be treated with care if they are likely to identify individuals. Fictitious examples are illustrated in Appendix A.

Data/information could be released internally to colleagues within Public Health Wales, externally to stakeholders or released to the public (including the media).

Where cases have been reported by the press and named, the details are in the public domain but a press report is not the same as confirmation from an official source. This would not mean that PHW can then release this information.

Relevant Public Health Wales staff who provide data/information will advise of any restrictions of use and, if necessary contact the relevant data supplier for advice or authorisation on the release of the data/information. Where there is risk of disclosure there needs to be consideration of how the data/information could be disclosed. The following approaches should be considered:

1. the design of the table
2. modifying the values in the table
3. adjusting the data/information before tables are designed

Possible ways to maintain the usefulness of the data/information, while avoiding small numbers are:

- use of a statement, e.g. over the ten year period there were on average 1.2 cases per year
- Restructuring tables by combining categories in which small number counts appear (e.g. age groups, time periods, areas)
- Suppression of cells with small numbers, or cells which, when taken with other data/information in the table, lead to small numbers counts with secondary suppression

In some circumstances there may be legitimate reasons to release uncontrolled data, e.g. in the public health assessment of potential disease clusters by health professionals.

If there is any uncertainty the appropriate Director or Team Lead should be approached for advice and approval. The Public Health Wales Caldicott Guardian is the Executive Director of Public Health Services and provides advice on release of potentially identifiable information owned by Public Health Wales. The accountable Executive Director for Information is the Senior Information Risk Owner – Quality, Nursing & Other Health Care Professionals.

Public Health Wales has no degree of control on the use of the data/information once released to the requestor. This is why terms and conditions are often stipulated. If there remains a doubt the data/information should not be released. If a breach occurs it must be logged.

In some instances it will not be possible to release the data/information requested. In this instance an individual within Public Health Wales will contact the requestor and explain the reason why the request was refused. If required this can be escalated to a senior member of staff. Decisions for non-release would be documented.

## **2.2 Data owned by others**

### **a) Data owned and controlled by another provider**

If the data is owned and controlled by another provider, PHW must abide by contracts, protocols, guidance and data sharing agreements which place restrictions on the use and disclosure of data/information.

One example would be the Data Access Agreements it has with the Office for National Statistics. These place restrictions on the release of statistics where harm or distress could be caused to an individual, or in the case of death records, to the living relatives of those who have died. This could be by releasing small counts at low geographies (sub national).

The Welsh Health Survey (WHS) owned by WG is designed to provide data at national and local authority level and can also be analysed at local health board level. The main source of Welsh Health Survey information is the [Welsh Assembly Government website](#)<sup>1</sup>.

- Use of the information should comply both with the provisions of the Data Protection Act 1998 and the [Code of Practice for Official Statistics: Principles 5 - Confidentiality and 8 - Frankness and accessibility](#)<sup>2</sup>.

Information should not be passed to any organisation without obtaining prior permission if required.

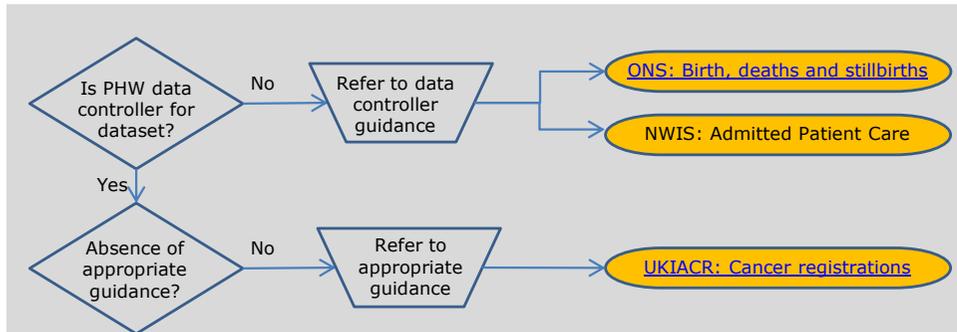
### **b) Data owned by Public Health Wales**

Primary data collected by Public Health Wales would need to adhere to the guidance and flowchart in this document.

### **2.3 Assessing information governance concerns**

The following flow chart has been developed to direct Public Health Wales staff to pre-existing guidance or to provide a framework for Public Health Wales data controllers to have developed flowcharts for their own purposes.

Details of how the linked policy is to be accomplished according to staff roles and responsibilities as set out above. This section provides direction for the users of the policy/procedure document and should be written in a clear and concise way in a framework that outlines a logical, progressive process.

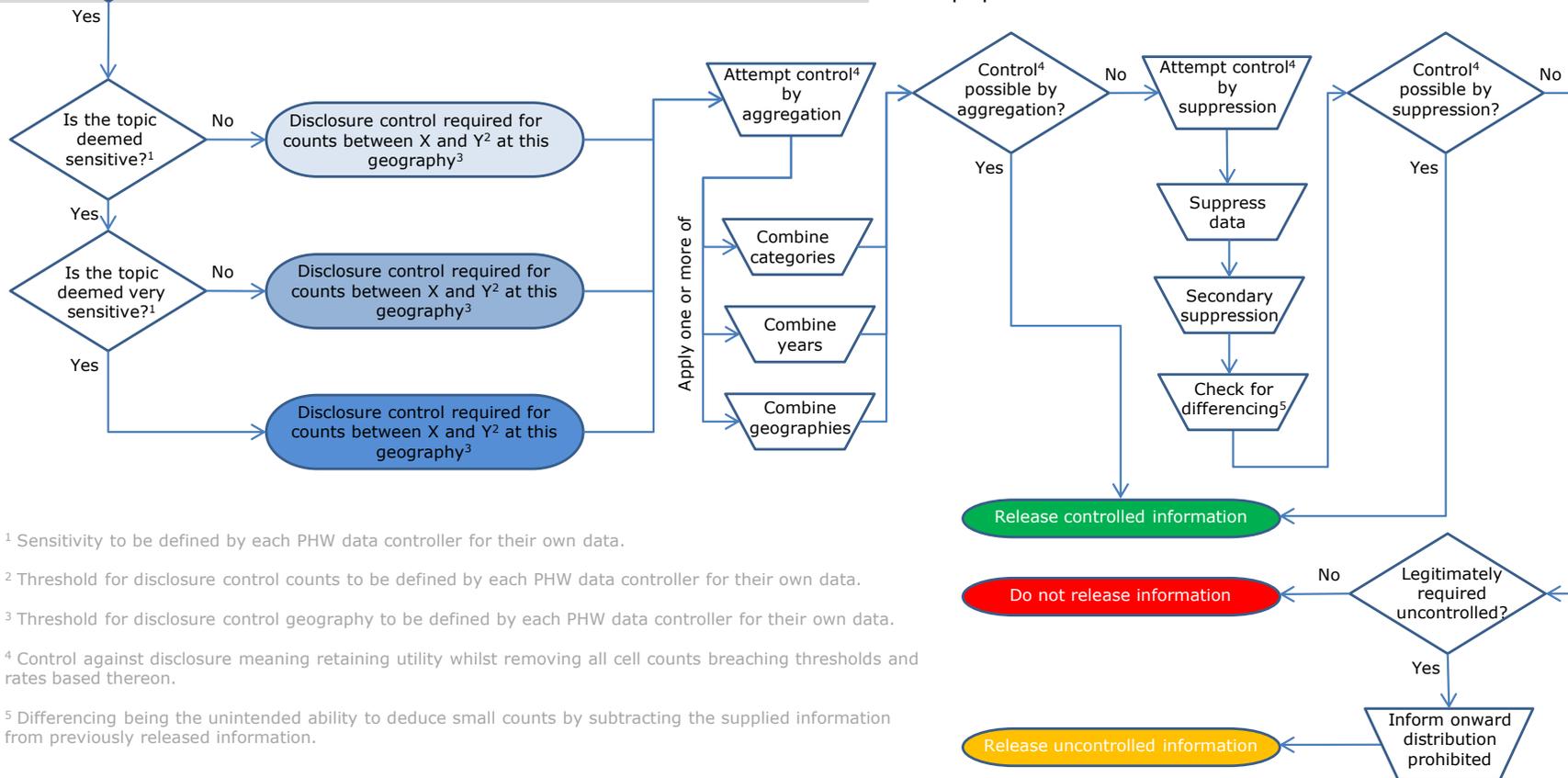


### Disclosure control: Guidance framework

**Purpose:** To enable Public Health Wales programmes and services to develop dataset specific guidance.

**Aim:** To ensure Public Health Wales meets the needs of requestors whilst ensuring it does not unintentionally disclose identifiable information in any published or supplied tabulations or images.

**Use:** Data controllers for datasets that fall outside of the grey shaded area should modify the remainder of the flow chart to develop specific rules for their datasets.



<sup>1</sup> Sensitivity to be defined by each PHW data controller for their own data.

<sup>2</sup> Threshold for disclosure control counts to be defined by each PHW data controller for their own data.

<sup>3</sup> Threshold for disclosure control geography to be defined by each PHW data controller for their own data.

<sup>4</sup> Control against disclosure meaning retaining utility whilst removing all cell counts breaching thresholds and rates based thereon.

<sup>5</sup> Differencing being the unintended ability to deduce small counts by subtracting the supplied information from previously released information.

### **3 What to do if there is a breach (incident reporting)**

A breach could be described as the release of small number at low geographical area; data/information being passed to others who release the data/information to others who should not have access to it.

Where it is known that there are breaches of information governance principles/rules or official statistics code of practice, the process for logging a risk or an incident is via the Public Health Wales [DATIX](#) Management system, used to report all Public Health Wales incidents.

Ways of minimising risk of such an incident to still meet the expectations and needs of the requestor may include the use of table redesign to aggregate small cell counts across categories or over time to maintain the usefulness of the data/information. The issue of data confidentiality should be discussed with the requestor and appropriately documented.

## 4 Appendices

### Appendix A

#### Example 1 – denominator smaller than 3,000 (ONS threshold for entire population)

An individual with a special interest in conception statistics could discover from a table that only a small number of very young women have conceived in a particular local area. The small number in the cell doesn't tell the intruder who the women are but it may prompt them to follow up other sources of information to locate the individuals and discover – and disclose – more details.

#### Example 2 – indirect disclosure

A table of statistics for psychiatric services at a hospital shows admissions by single years of age, and diagnosis. Attribute disclosure has occurred if someone, who knows their neighbour was admitted for such a service, discovers from the small number statistic that they have a diagnosis of schizophrenia.

#### Example 3 – sensitive issue and indirect disclosure through cross-tabulation (denominator smaller than 3,000)

In the following table, because of the six and the zeros in the “Indian, Pakistani or Bangladeshi” column, if you know someone in this ethnic group who has the infection then you also know how they were infected.

**Table 1: Infection route by ethnic group**

|                                  | White      | Black<br>(Caribbean,<br>Asian or<br>other) | Indian,<br>Pakistani or<br>Bangladeshi | Other,<br>mixed |
|----------------------------------|------------|--|--|-----------------|
| <b>Sex between men</b>           | 86         | 10   | 6                                      | 8               |
| <b>Sex between men and women</b> | 10         | 1  | 0                                      | 2               |
| <b>Injecting drug use</b>        | 8          | 1  | 0                                      | 1               |
| <b>Other</b>                     | 4          | 0  | 0                                      | 1               |
| <b>Total</b>                     | <b>108</b> | <b>12</b>                                  | <b>6</b>                               | <b>12</b>       |

A solution could be combining columns i.e. White, Black, Asian/other

#### Example 4 – self disclosure causing distress

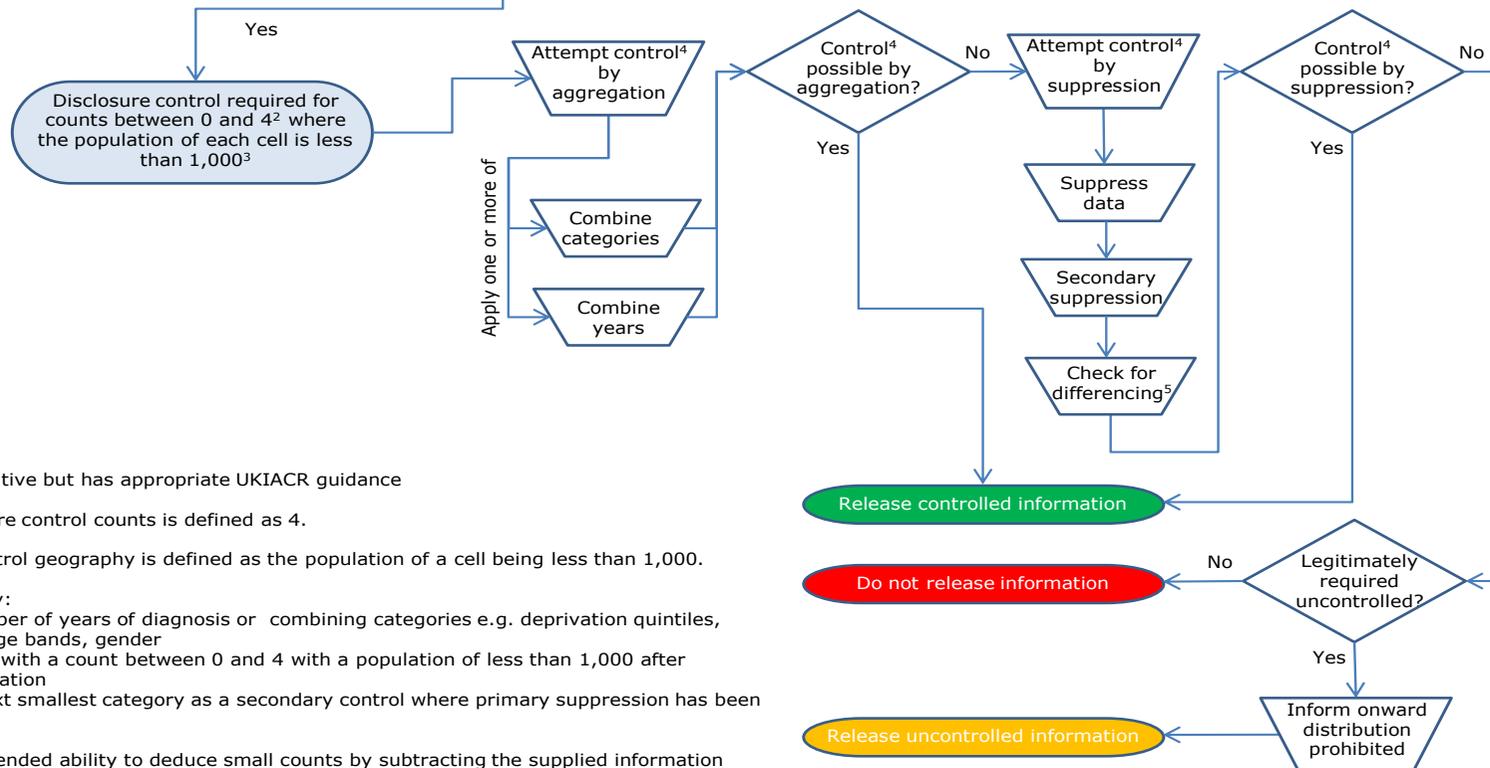
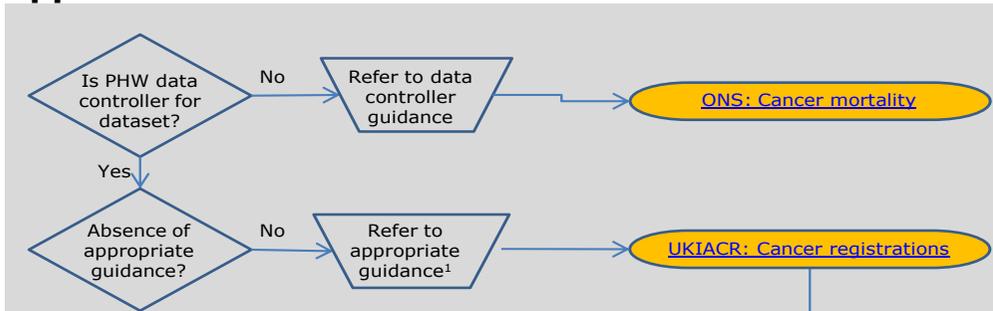
A statistic showing attendance at a drug misuse clinic by age and sex has a count of 1 for a particular ward. The individual may in fact be the only person who knows who this 1 is but they may feel exposed by the statistic. If this fear is communicated to their peers, it may spread, and the result may be a lack of trust in the confidentiality of their use of the clinic.

## Appendix B – Reference documents

1. Welsh Government. 2011. Welsh health survey. [Online]. Available at: <http://wales.gov.uk/topics/statistics/theme/health/health-survey/;jsessionid=6fpQTx2dzsh8HyckhltXYGCzV6nwWNdh6NW7ph5Dp31RxH5bQfvj!-1883660344?lang=en> [Accessed: 8 November 2011]
2. Office for National statistics. *Review of the dissemination of health statistics: Confidentiality guidance*. Newport: ONS; 2006. Available at: <http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-of-health-statistics/confidentiality-guidance.pdf> [Accessed: 2 November 2011]
3. NWIS. *Statistical Disclosure Control protocol*. (To be finalised.)
4. Review of the Dissemination of Health Statistics: Working Paper 5: References and other Guidance. Newport: ONS; 2006. Available at: <http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-of-health-statistics/working-paper-5--references.pdf> [Accessed: 2 November 2011]
5. Information Commissioners Office. FOI Policy Knowledge Base. Section 40. [Online] Available at: <https://ico.org.uk/media/1213/personal-information-section-40-and-regulation-13-foia-and-eir-guidance.pdf>
6. Office for National Statistics. Briefing Note: ONS policy on protecting confidentiality within birth and death statistics. Newport: ONS; 2010 Available at: <http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-birth-and-death-statistics/index.html> [Accessed: 2 November 2011]
7. BINOCAR. The BINOCAR Small Numbers Policy. Newcastle Upon Tyne: BINOCAR; 2011. Available at: [www.binocar.org/content/BINOCAR%20Small%20Numbers%20Policy.docx](http://www.binocar.org/content/BINOCAR%20Small%20Numbers%20Policy.docx) [Accessed: 8 November 2011]
8. Health and Social Care Act 2012. Chapter 14. Section 251. London: TSO; 2012. Available at: [http://www.legislation.gov.uk/ukpga/2012/7/pdfs/ukpga\\_20120007\\_en.pdf](http://www.legislation.gov.uk/ukpga/2012/7/pdfs/ukpga_20120007_en.pdf) [Accessed 18 February 2014]

9. UKACR guidelines on release of: a) individual level anonymised information and b) tabular information based on small populations or small cell counts (potentially identifiable information). Available at:  
<http://www.google.co.uk/url?url=http://www.ncin.org.uk/view%3Frid%3D1705&rct=j&frm=1&q=&esrc=s&sa=U&ved=0ahUKEwjew5WPYnNAhVIC8AKHXV6ASoQFggZMAE&usg=AFQjCNFpxFbH8sFYhRJXpXIwzur38ua-Ng>
10. Public Health (Control of Diseases) Act 1984. Chapter 22. London: HMSO; 1984 Available at:  
<http://www.legislation.gov.uk/ukpga/1984/22> [Accessed: 2 November 2011]
11. Health Protection (Notification) (Wales) Regulations 2010. London: TSO; 2010. Available at:  
<http://www.legislation.gov.uk/wsi/2010/1546/contents/made> [Accessed: 2 November 2011]
12. UK Statistics Authority. Code of practice for official statistics. January 2009. London: UK Statistics Authority; 2009. Available at: <https://www.statisticsauthority.gov.uk/monitoring-and-assessment/code-of-practice/> [Accessed: 2 November 2011]
13. Anonymisation: managing data/information protection risk code of practice (2012). Available at:  
<https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/>

# Appendix C



## Disclosure control: Guidance for publishing cancer incidence

**Aim:** To ensure the Welsh Cancer Intelligence and Surveillance Unit meets the needs of stakeholders whilst ensuring it does not unintentionally disclose identifiable information in its annual reports and associated products.

<sup>1</sup> Cancer is deemed very sensitive but has appropriate UKIACR guidance

<sup>2</sup> Upper threshold for disclosure control counts is defined as 4.

<sup>3</sup> Threshold for disclosure control geography is defined as the population of a cell being less than 1,000.

<sup>4</sup> Control against disclosure by:

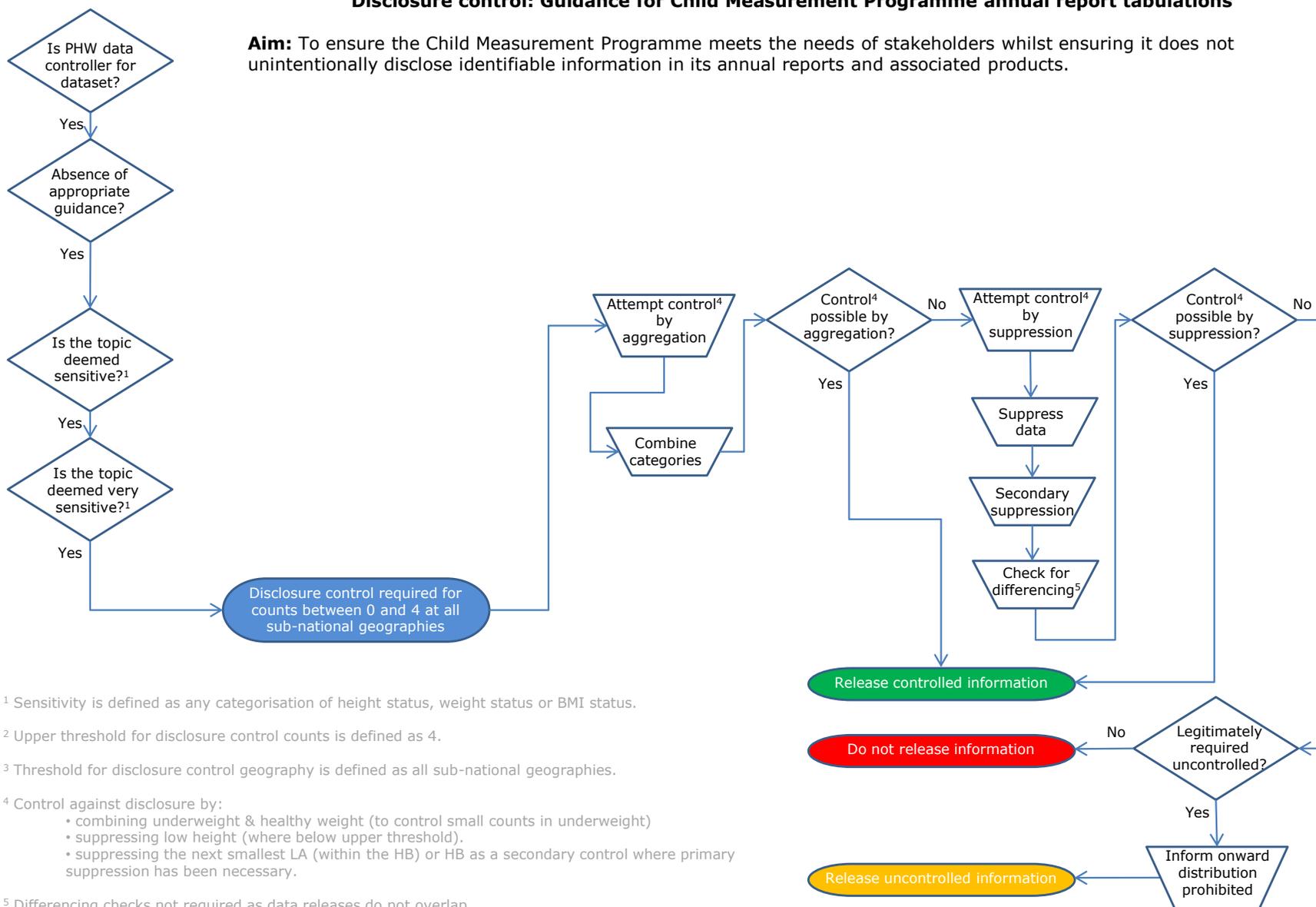
- combining the number of years of diagnosis or combining categories e.g. deprivation quintiles, stage at diagnosis, age bands, gender
- suppressing counts with a count between 0 and 4 with a population of less than 1,000 after combining by aggregation
- suppressing the next smallest category as a secondary control where primary suppression has been necessary

<sup>5</sup> Differencing being the unintended ability to deduce small counts by subtracting the supplied information from previously released information.

## Appendix D

### Disclosure control: Guidance for Child Measurement Programme annual report tabulations

**Aim:** To ensure the Child Measurement Programme meets the needs of stakeholders whilst ensuring it does not unintentionally disclose identifiable information in its annual reports and associated products.



<sup>1</sup> Sensitivity is defined as any categorisation of height status, weight status or BMI status.

<sup>2</sup> Upper threshold for disclosure control counts is defined as 4.

<sup>3</sup> Threshold for disclosure control geography is defined as all sub-national geographies.

<sup>4</sup> Control against disclosure by:

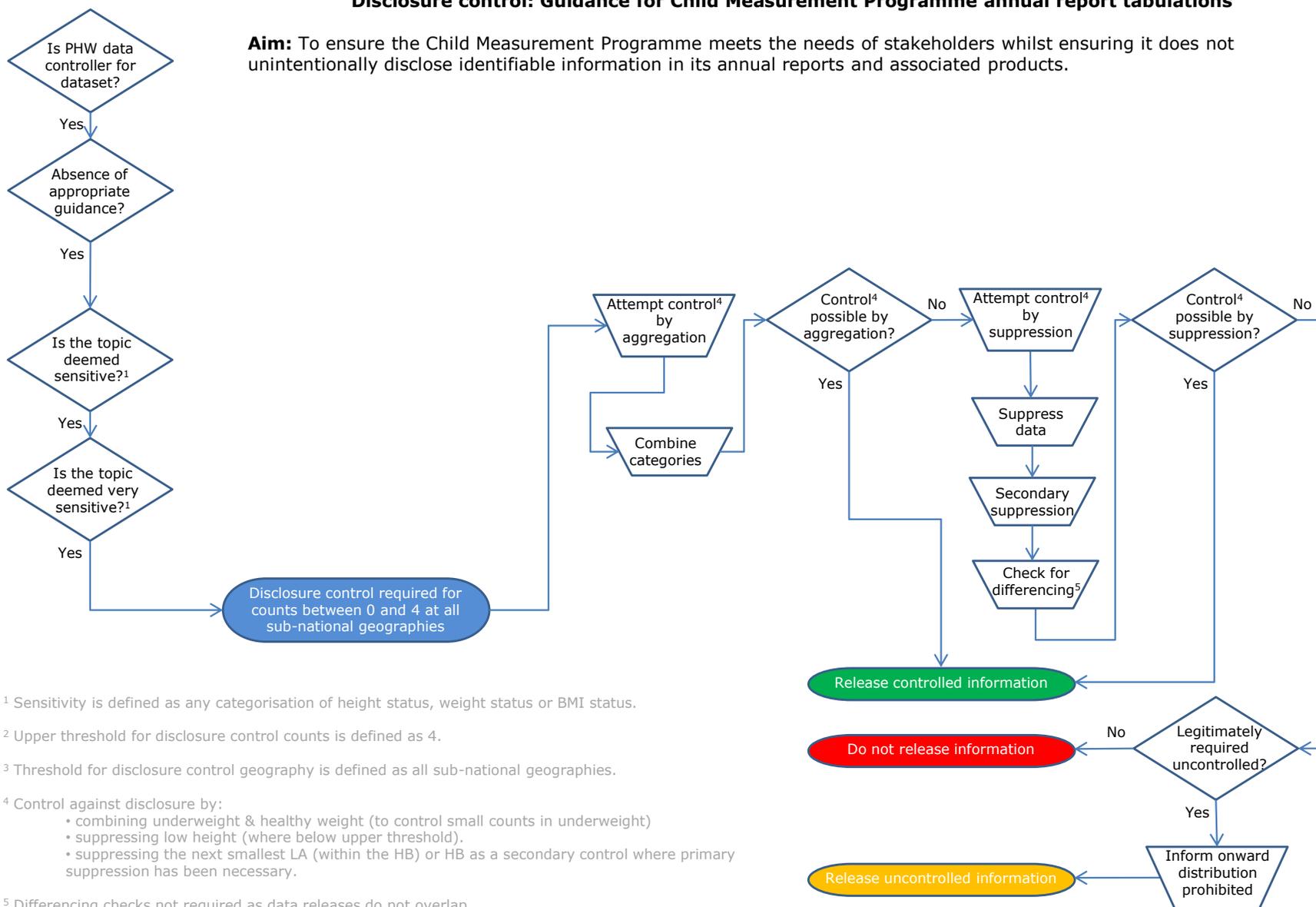
- combining underweight & healthy weight (to control small counts in underweight)
- suppressing low height (where below upper threshold).
- suppressing the next smallest LA (within the HB) or HB as a secondary control where primary suppression has been necessary.

<sup>5</sup> Differencing checks not required as data releases do not overlap.

## Appendix E

### Disclosure control: Guidance for Child Measurement Programme annual report tabulations

**Aim:** To ensure the Child Measurement Programme meets the needs of stakeholders whilst ensuring it does not unintentionally disclose identifiable information in its annual reports and associated products.



<sup>1</sup> Sensitivity is defined as any categorisation of height status, weight status or BMI status.

<sup>2</sup> Upper threshold for disclosure control counts is defined as 4.

<sup>3</sup> Threshold for disclosure control geography is defined as all sub-national geographies.

<sup>4</sup> Control against disclosure by:

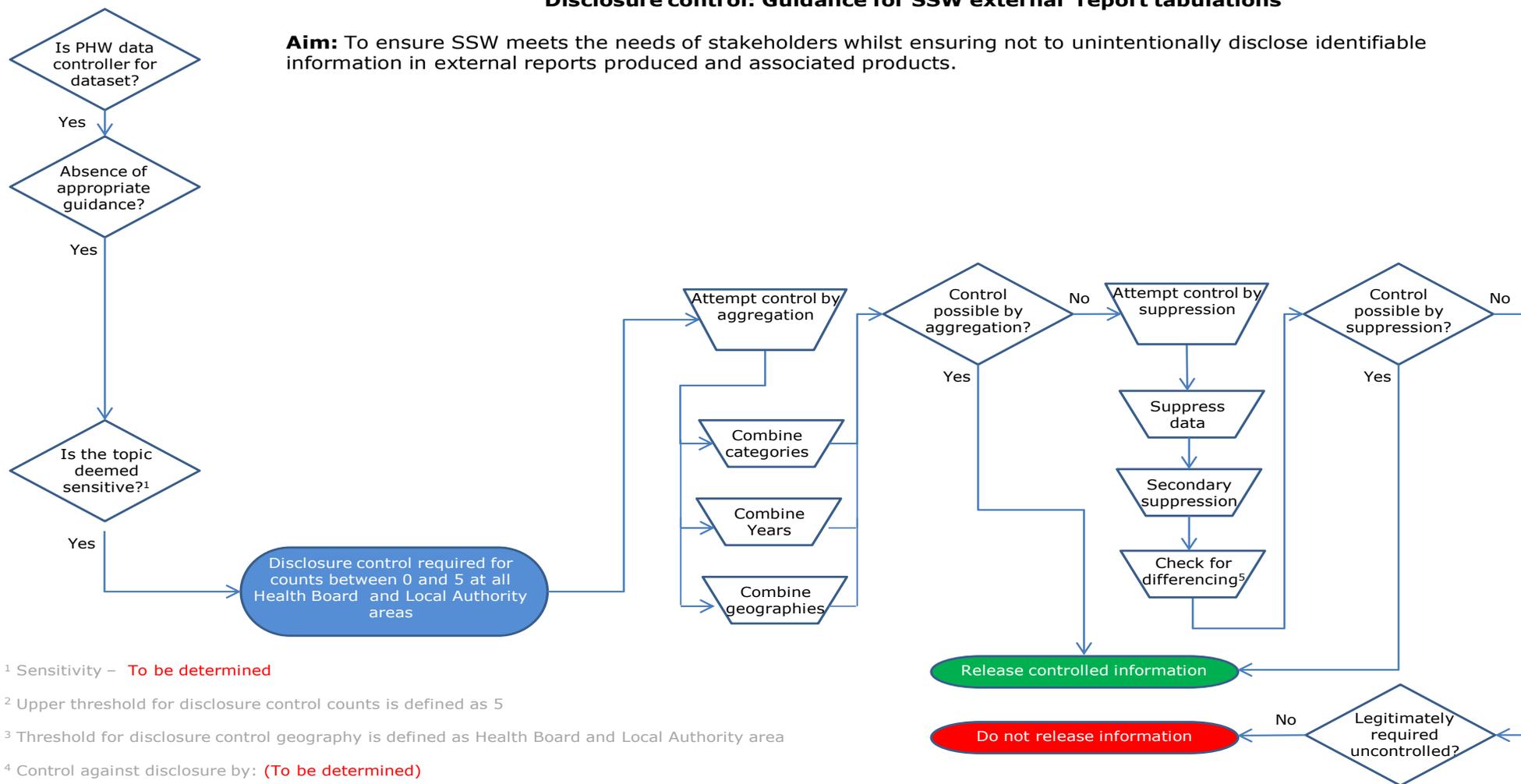
- combining underweight & healthy weight (to control small counts in underweight)
- suppressing low height (where below upper threshold).
- suppressing the next smallest LA (within the HB) or HB as a secondary control where primary suppression has been necessary.

<sup>5</sup> Differencing checks not required as data releases do not overlap.

## Appendix F

### Disclosure control: Guidance for SSW external report tabulations

**Aim:** To ensure SSW meets the needs of stakeholders whilst ensuring not to unintentionally disclose identifiable information in external reports produced and associated products.



<sup>1</sup> Sensitivity – **To be determined**

<sup>2</sup> Upper threshold for disclosure control counts is defined as 5

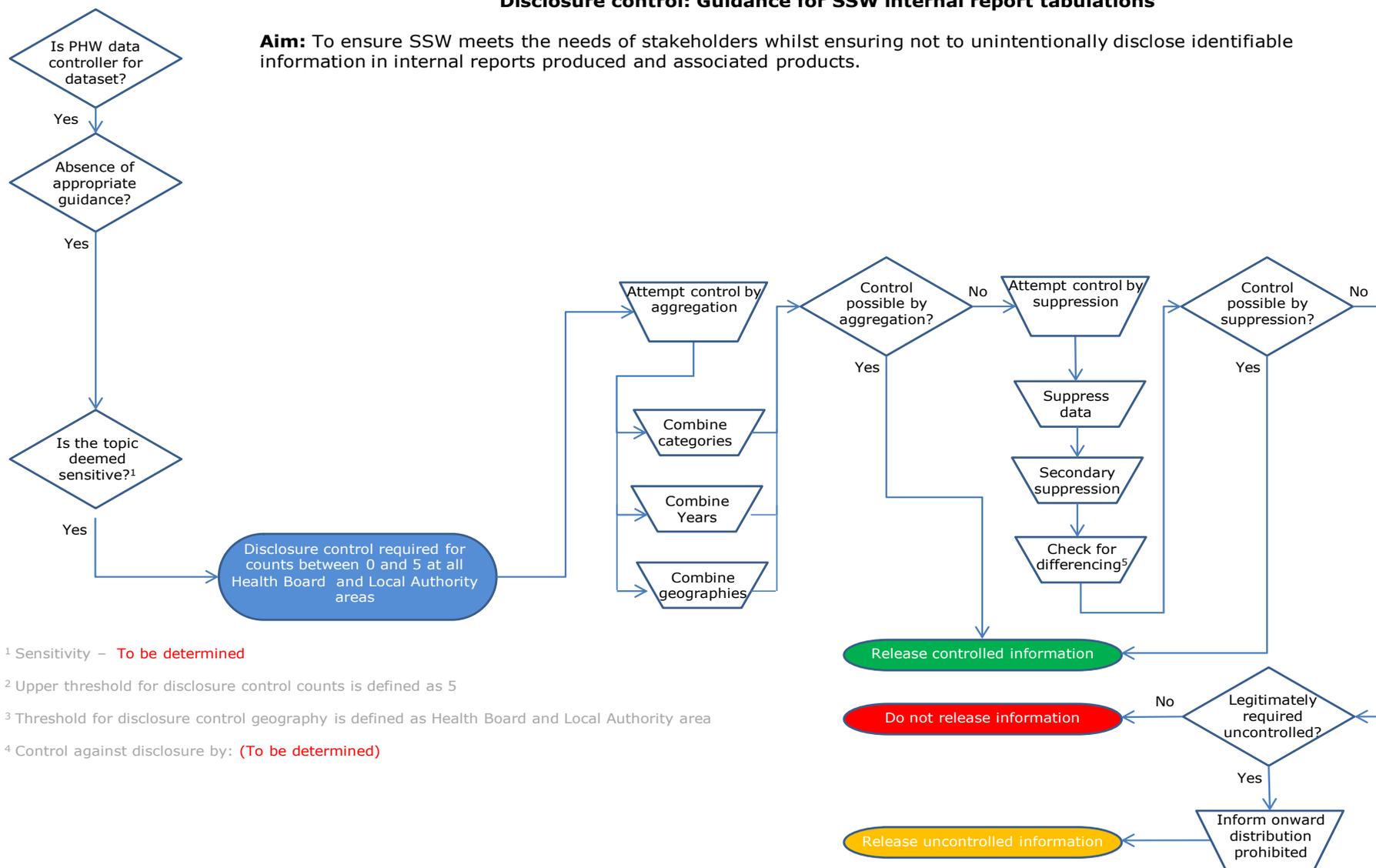
<sup>3</sup> Threshold for disclosure control geography is defined as Health Board and Local Authority area

<sup>4</sup> Control against disclosure by: **(To be determined)**

## Appendix G

### Disclosure control: Guidance for SSW internal report tabulations

**Aim:** To ensure SSW meets the needs of stakeholders whilst ensuring not to unintentionally disclose identifiable information in internal reports produced and associated products.



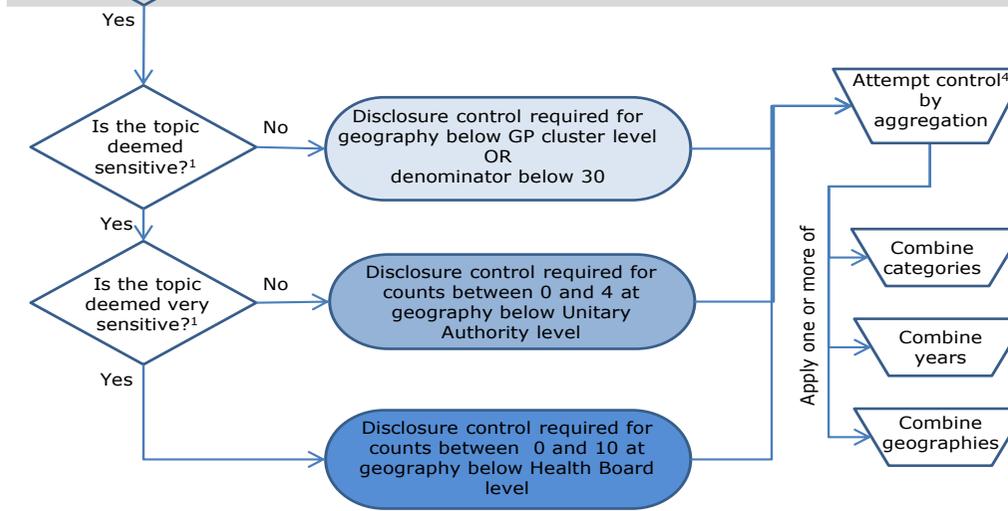
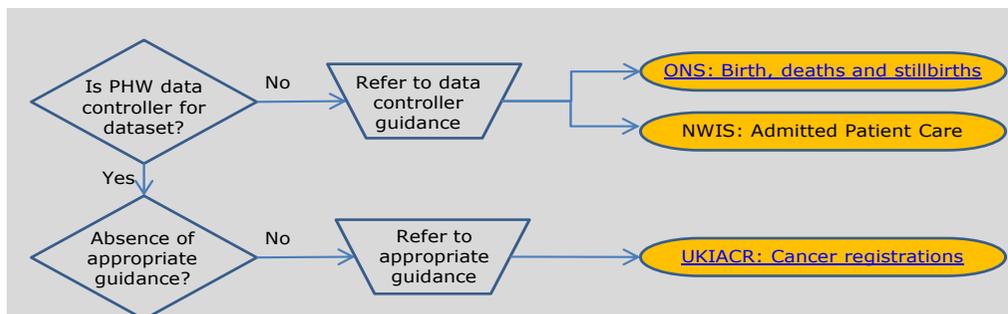
<sup>1</sup> Sensitivity – To be determined

<sup>2</sup> Upper threshold for disclosure control counts is defined as 5

<sup>3</sup> Threshold for disclosure control geography is defined as Health Board and Local Authority area

<sup>4</sup> Control against disclosure by: (To be determined)

## Appendix H



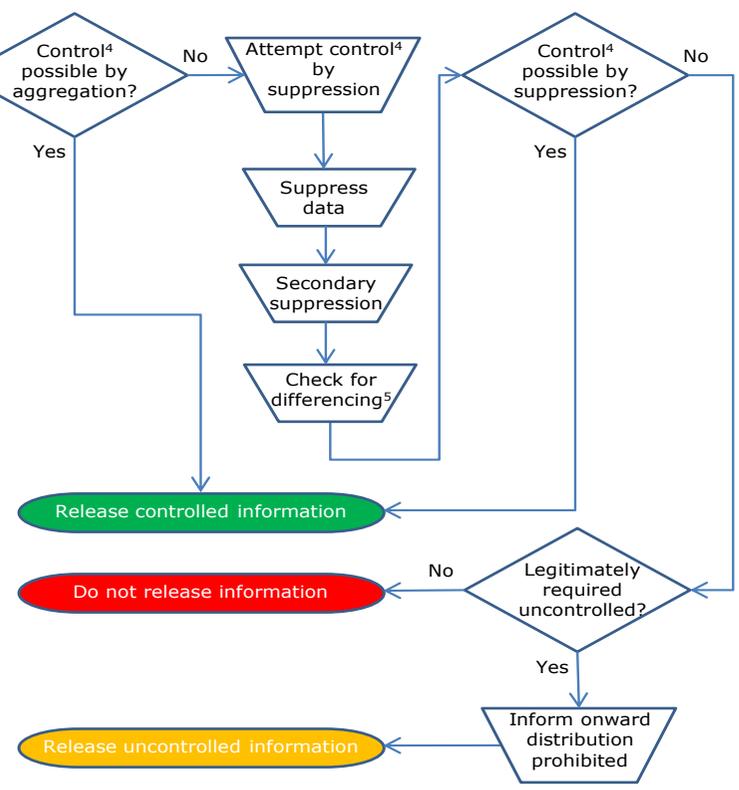
1 Sensitivity to be defined by each PHW data controller for their own data.  
 2 Threshold for disclosure control counts to be defined by each PHW data controller for their own data.  
 3 Threshold for disclosure control geography to be defined by each PHW data controller for their own data.  
 4 Control against disclosure meaning retaining utility whilst removing all cell counts breaching thresholds and rates based thereon.  
 5 Differencing being the unintended ability to deduce small counts by subtracting the supplied information from previously released information.

**Screening Disclosure control: Guidance framework**

**Purpose:** To enable Public Health Wales programmes and services to develop dataset specific guidance.

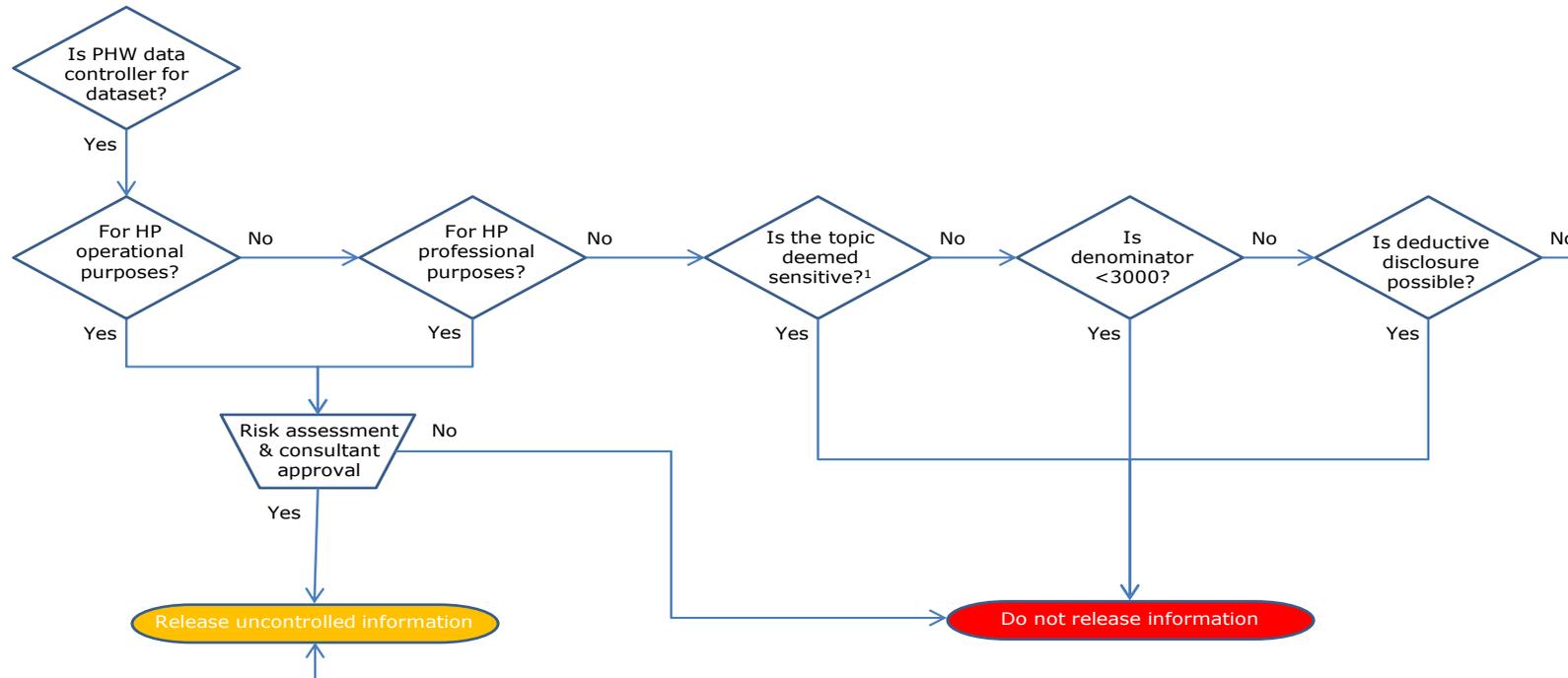
**Aim:** To ensure Public Health Wales meets the needs of requestors whilst ensuring it does not unintentionally disclose identifiable information in any published or supplied tabulations or images.

**Use:** Data controllers for datasets that fall outside of the grey shaded area should modify the remainder of the flow chart to develop specific rules for their datasets.



### Disclosure control: Guidance for Health Protection

**Aim:** To ensure Health Protection meets the needs of stakeholders whilst ensuring it does not unintentionally disclose identifiable information.



<sup>1</sup> Sensitivity is defined as including but not limited to sexually transmitted and blood borne infections, single cases of emerging infections such as Ebola or Zika, or data that could be used or perceived to be used for performance management purposes.