

Appendix 2

Troubleshooting the differences between HESES04 and the HESES04 re-creation

Purpose

1. This appendix aims to help institutions identify the cause of any discrepancies between their HESA student data and HESES04 return. It is expected that institutions will have worked through this appendix and consulted the web-based FAQ page on the HEFCE website (under Learning & teaching/Data collection/Statistics derived from HESA data for monitoring and allocation of funding) before seeking assistance from HEFCE on resolving discrepancies.

Using this appendix

2. Figure 1 provides a systematic method for identifying at what point discrepancies between the returns occur. The subsequent paragraphs give possible causes for each discrepancy. These causes can be grouped into two categories:

- errors in completing specific fields on the HESA return (addressed in this appendix)
- problems of fit with the HESES04 re-creation algorithms (addressed in Appendix 3).

3. The diagnostic diagram in Figure 1 can be used to help identify errors in completing specific fields on the HESA return.

4. The match between HESES and HESA data is unlikely to be exact, due to estimates made when returning HESES and approximations made in the re-creation algorithms (see Appendix 3 for further details). Therefore, when using the diagnostic diagram we expect institutions to exercise their own judgement to decide when small differences between the two data sources are not significant. However, institutions need to be aware that small differences may accumulate and become significant. When the cause of a significant difference cannot be determined, it may be necessary to backtrack to find the root of the problem.

5. The description in this appendix is aimed at identifying weaknesses in the HESA data rather than HESES. Therefore it is possible to follow the diagnostic diagram given in Figure 1 without resolving the discrepancies if they are due to errors in HESES04.

6. Throughout this appendix, fields taken from the HESA return or derived as part of the re-creation are shown in capitals using the names given in Tables 7 and 8 of Appendix 1.

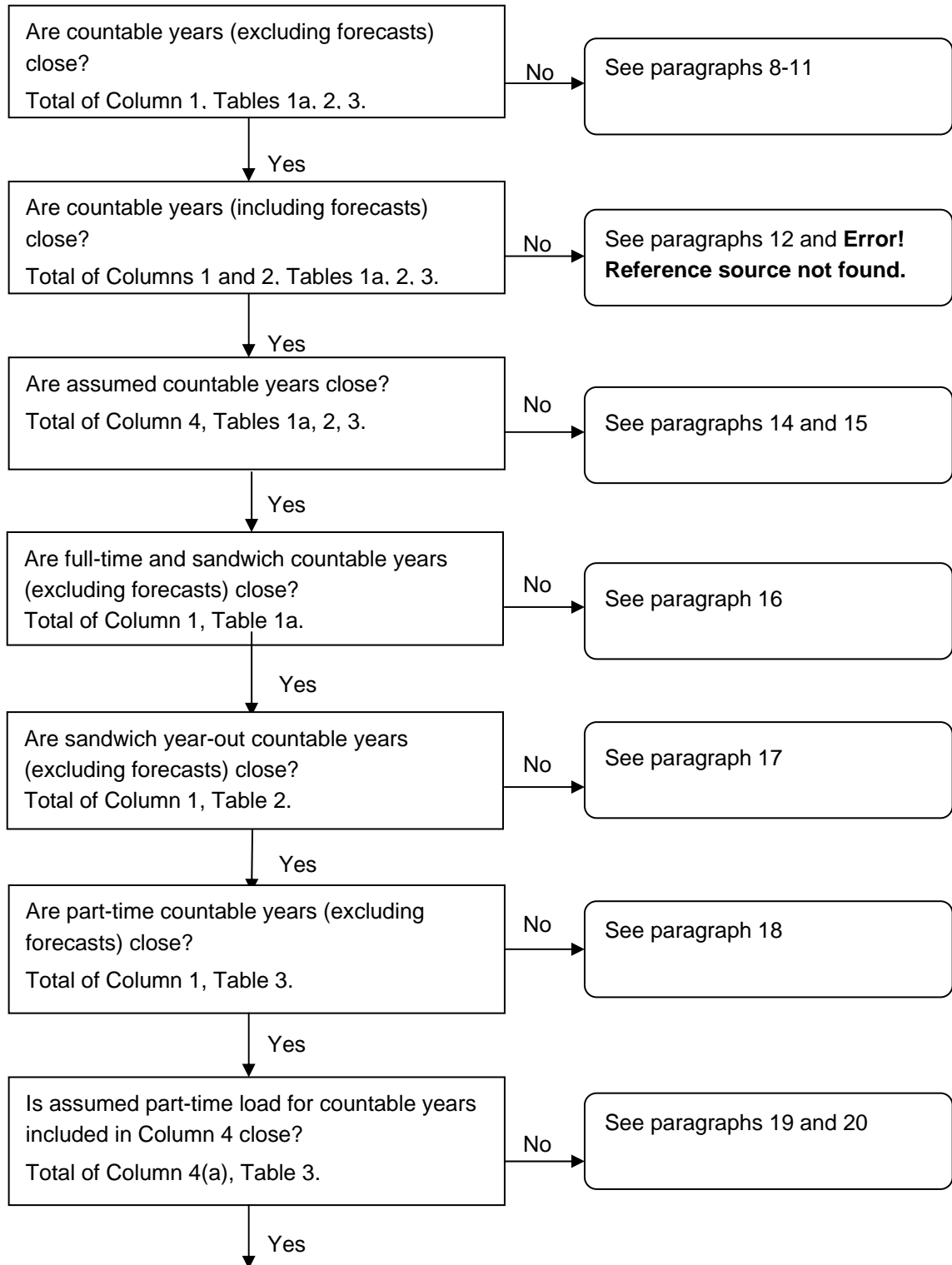
Using the individualised file

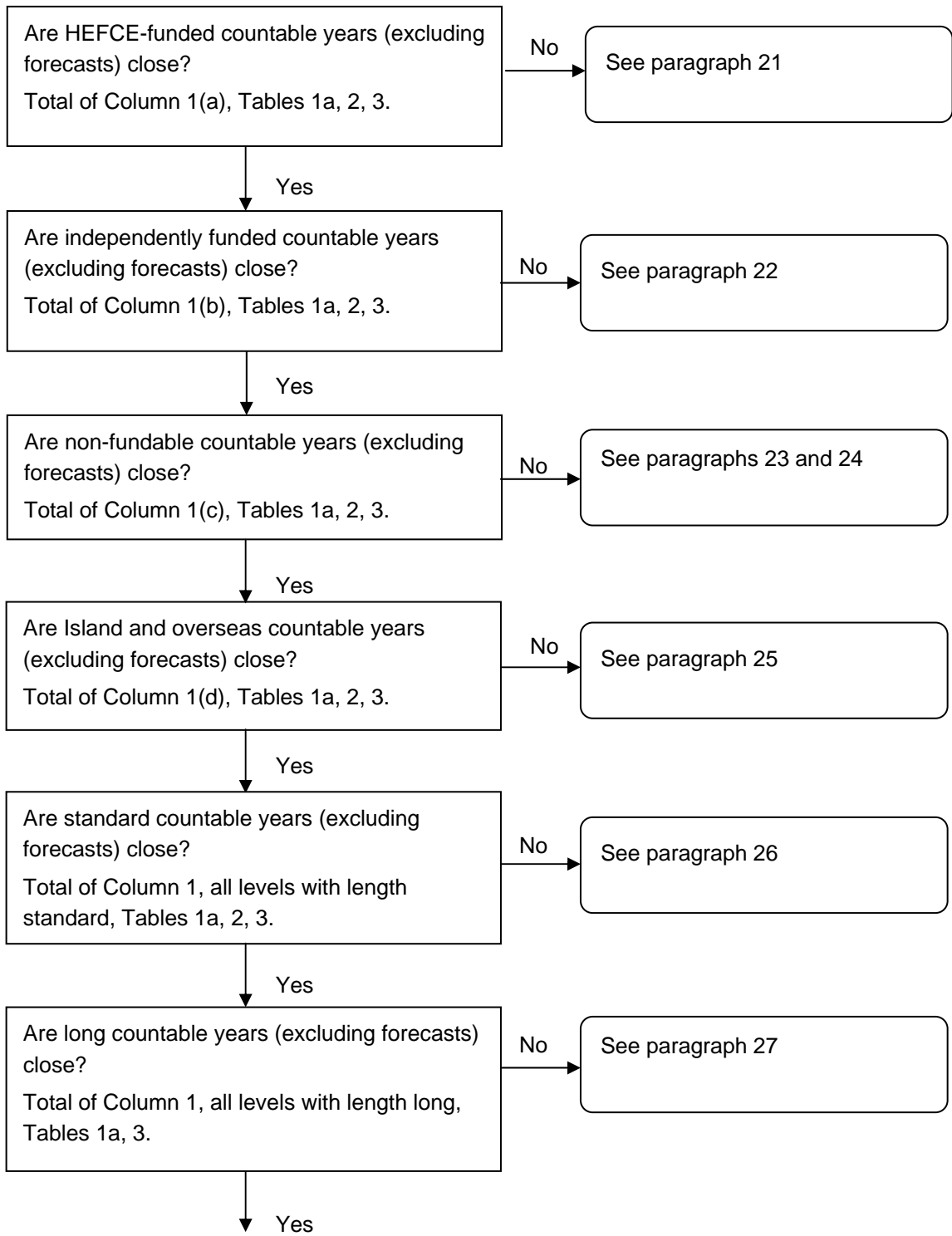
7. The individualised files show the allocation of students to cells within the equivalent re-creation tables and, where relevant, details of why they were excluded. For institutions who use Microsoft Excel with individualised files that do not contain greater than 65,530

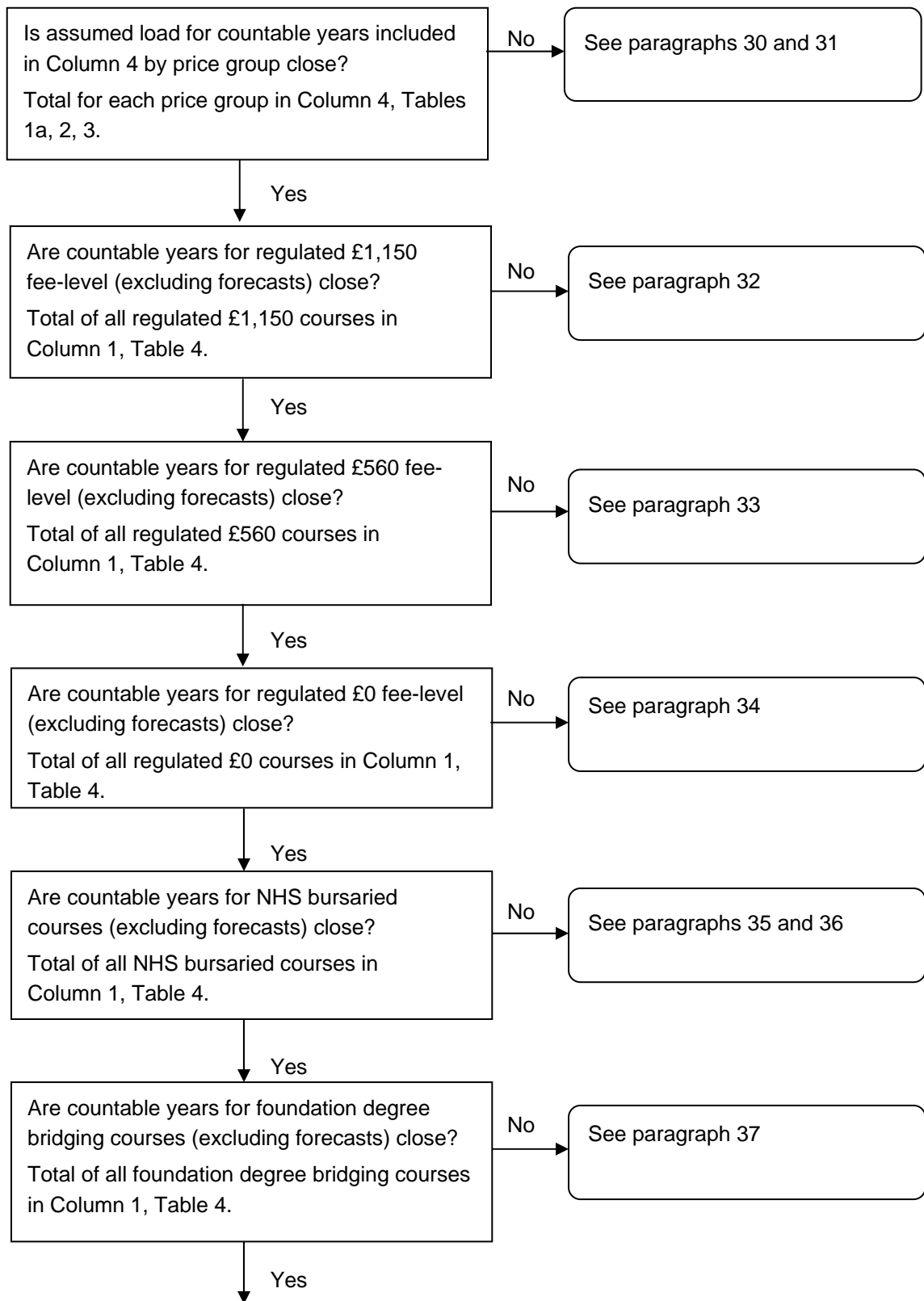
records, the following guidance will assist them in the troubleshooting process. All records with a specific value in a specific field can be found in the following way:

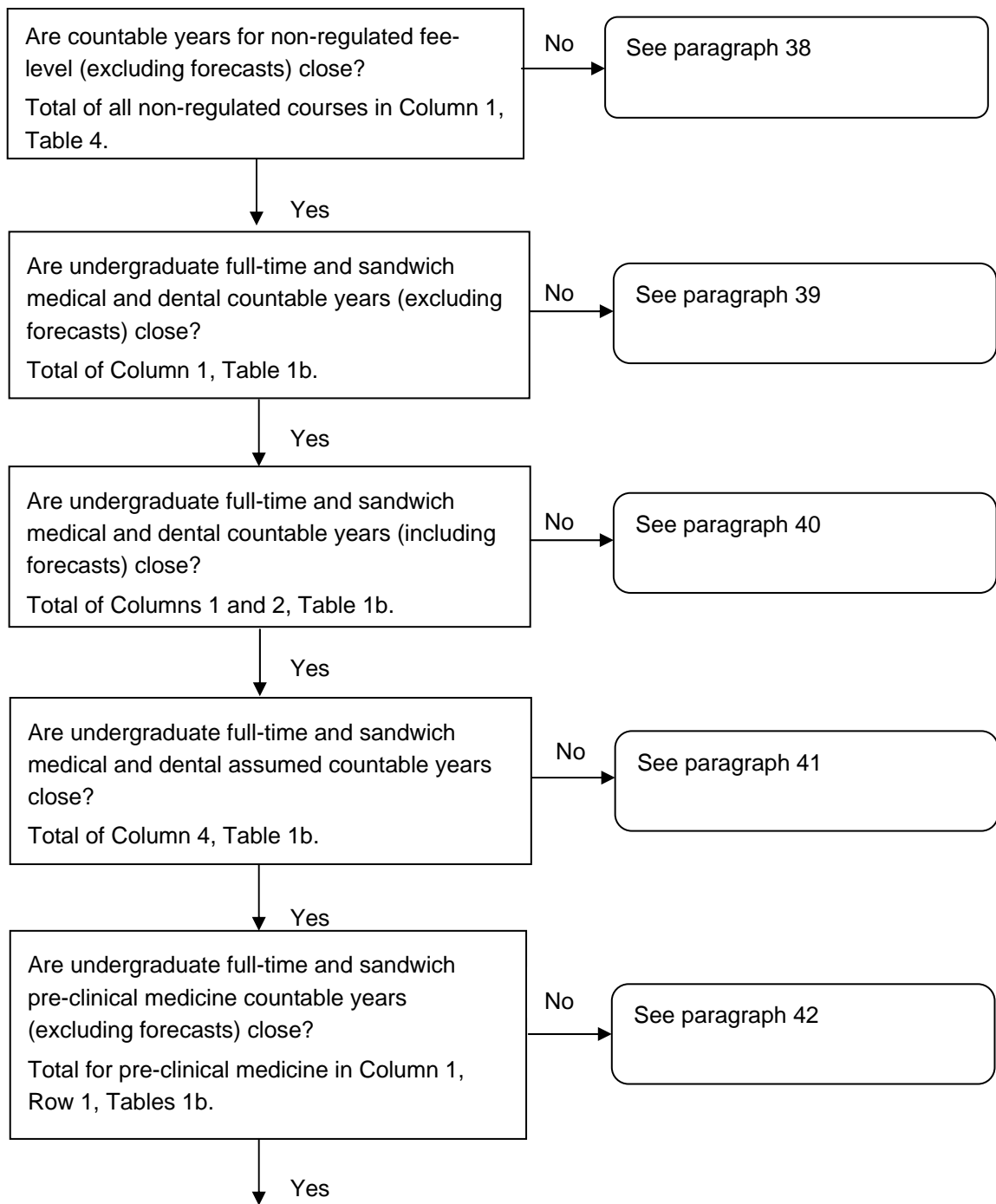
- a. In order to open an individualised file, go into Microsoft Excel and click <File>, <Open>. You will need to specify 'All Files' in the 'Files of type' box before the individualised file will appear in the file listing. Once you have selected the file, the 'Text Import Wizard' will appear. Ensure that 'Delimited' is selected near the top of the window, then click 'Next'. On the next page, uncheck 'Tab' and check 'Comma'. Click 'Finish' to open the file.
- b. Select the row containing the field headings.
- c. Select <Filter> from the <Data> menu and then <Autofilter>.
- d. Click on the arrow in the column containing the data which you want to filter.
- e. Either select a specific value or select <Custom> to apply a comparison operator other than equality.
- f. To select records using multiple fields, repeat steps d and e for each field.

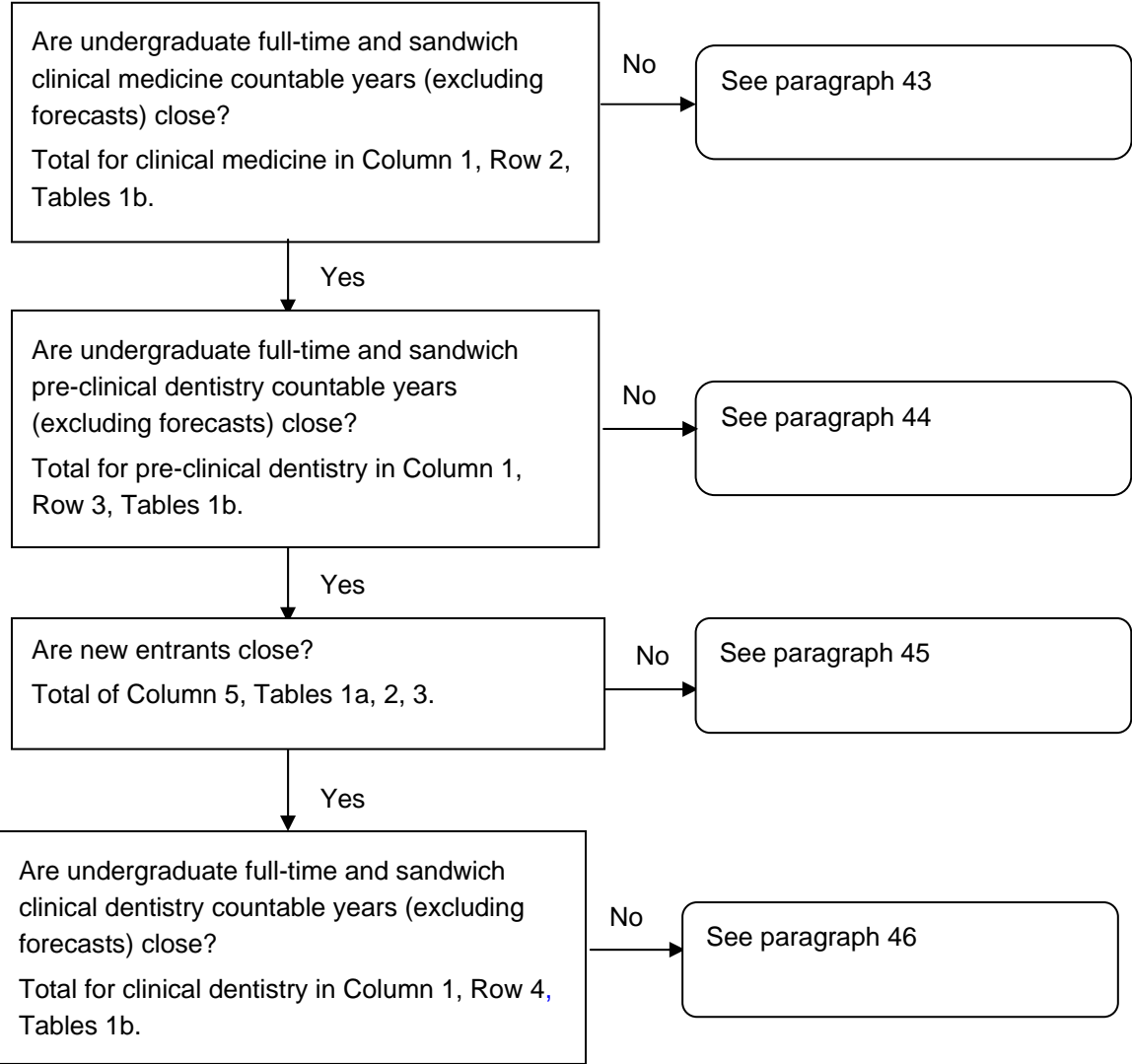
Figure 1 **Diagnostic flowchart**











Countable years (excluding forecasts)

8. To identify countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1. The algorithms for deriving HESEXCL and HESREG are given in paragraphs 68-69 and 65 respectively of Appendix 1.

9. Exceptionally a student generates two countable years of programme of study on the HESES return (see paragraph 11 of Annex D of 'HESES04: Higher Education Students Early Statistics Survey 2004-05', HEFCE 2004/31, for further details). The field for identifying multiple years of programme of study is STUBID. The algorithms for deriving STUBID are given in paragraphs 32-35 of Appendix 1.

10. If our record of the method used to return FTE on the HESA student record is incorrect, differences are likely to arise (see paragraphs 7-13 of Annex D of 'HESES04: Higher Education Students Early Statistics Survey 2004-05', HEFCE 2004/31, for further details). This is likely to cause a discrepancy between the two data sources for records where any of the following are true:

- FTE_TYPE = 1 and TYPEYR \neq 1
- FTE_TYPE = 2 and STULOAD = 0
- FTE_TYPE = 3 and COMDATE > 31 July 2004 and STULOAD = 0
- FTE_TYPE = 4 and DATELEFT > 31 July 2004 and FUNDCOMP \neq 2 and STULOAD = 0

11. We make an assumption about students writing up a thesis or a similar piece of work. Details of this assumption are given in paragraph 12 of Appendix 3.

Countable years (including forecasts)

12. To identify countable years (including forecasts) from the individualised file select HESEXCL = 0. The algorithms for deriving HESEXCL are given in paragraphs 68 and 69 of Appendix 1.

13. There may be a variance as a result of forecasting countable years on HESES04.

Assumed countable years

14. To identify assumed countable years from the individualised file select HESCOL4 = 1. Forecast non-completions are identifiable by HESEXCL = 0 and HESCOMP = 3. The algorithms for deriving HESCOL4, HESEXCL and HESCOMP are given in paragraphs 80, 68-69 and 66 respectively of Appendix 1.

15. There may be a variance as a result of forecasting non-completions on HESES04.

Full-time and sandwich countable years (excluding forecasts)

16. To identify full-time and sandwich countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESMODE = FTS. The algorithms for deriving HESEXCL, HESREG and HESMODE are given in paragraphs 68-69, 65 and 22 respectively of Appendix 1.

Sandwich year-out countable years (excluding forecasts)

17. To identify sandwich year-out countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESMODE = SWOUT. The algorithms for deriving HESEXCL, HESREG and HESMODE are given in paragraphs 68-69, 65 and 22 respectively of Appendix 1.

Part-time countable years (excluding forecasts)

18. To identify part-time countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESMODE = PT. The algorithms for deriving HESEXCL, HESREG and HESMODE are given in paragraphs 68-69, 65 and 22 respectively of Appendix 1.

Assumed part-time load for countable years included in Column 4

19. To identify assumed part-time countable years included in Column 4 from the individualised file select HESCOL4 = 1 and HESMODE = PT. To obtain the load for these countable years sum the values of HESESFTE. The algorithms for deriving HESCOL4, HESMODE and HESESFTE are given in paragraphs 80, 22 and 49-50 respectively of Appendix 1.

20. We make an assumption about the load of part-time students. Details of this assumption are given in paragraphs 5-7 of Appendix 3.

HEFCE-funded countable years (excluding forecasts)

21. To identify HEFCE-funded countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESTYPE = HOMEF. The algorithms for deriving HESEXCL, HESREG and HESTYPE are given in paragraphs 68-69, 65 and 23 respectively of Appendix 1.

Independently funded countable years (excluding forecasts)

22. To identify independently funded countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESTYPE = HOMEIF. The algorithms for deriving HESEXCL, HESREG and HESTYPE are given in paragraphs 68-69, 65 and 23 respectively of Appendix 1.

Non-fundable countable years (excluding forecasts)

23. To identify non-fundable countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESTYPE = HOMENF. The algorithms for deriving HESEXCL, HESREG and HESTYPE are given in paragraphs 68-69, 65 and 23 respectively of Appendix 1.

24. We make an assumption about non-fundable students. Details of this assumption are given in paragraph 15 of Appendix 3.

Island and overseas countable years (excluding forecasts)

25. To identify Island and overseas countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESTYPE = ISOV. The algorithms for deriving HESEXCL, HESREG and HESTYPE are given in paragraphs 68-69, 65 and 23 respectively of Appendix 1.

Standard countable years (excluding forecasts)

26. To identify standard countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and LENGTH = S. The algorithms for deriving HESEXCL, HESREG and LENGTH are given in paragraphs 68-69, 65 and 39 respectively of Appendix 1.

Long countable years (excluding forecasts)

27. To identify long countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and LENGTH = L. The algorithms for deriving HESEXCL, HESREG and LENGTH are given in paragraphs 68-69, 65 and 39 respectively of Appendix 1.

Countable years (excluding forecasts) by price group

28. To identify countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1. To obtain the proportion of activity in each price group sum the values of each of the price group fields (PRGA, PRGB, PRGC, PRGD, PRGMEDIA, PRGITT, PRGINSET). The algorithms for deriving HESEXCL, HESREG and price groups are given in paragraphs 68-69, 65 and 54-58 respectively of Appendix 1.

29. We make assumptions about students on low-credit bearing courses. Details of this assumption are given in paragraphs 18 and 19 of Appendix 3.

Assumed load (FTE) for countable years included in Column 4 (and Column 4a for part-time) by price group

30. To identify assumed countable years included in Column 4 (and Column 4a for part-time) from the individualised file select HESCOL4 = 1. To obtain the load for these countable years multiply the values of each of the price group fields (PRGA, PRGB, PRGC, PRGD, PRGMEDIA, PRGITT, PRGINSET) with HESESFTE, sum the values and divide by 100. The algorithms for deriving HESCOL4, HESESFTE and price groups are given in paragraphs 80, 49-50 and 54-58 respectively of Appendix 1.

31. For the 'Summary comparison of price group activity between HESES04 re-creation and the HESES04' table, the assumed load (FTE) is calculated by summing all values in Column 4 for Table 1a, half the values in Column 4 for Table 2 and all values in Column 4a for Table 3.

Regulated £1,150 fee-level countable years (excluding forecasts)

32. To identify regulated £1,150 fee-level countable years (excluding forecasts) from the individualised file select HESEXCL = 0, HESREG = 1 and HESFEELV = 1150. The algorithms for deriving HESEXCL and HESFEELV are given in paragraphs 68-69 and 37 respectively of Appendix 1.

Regulated £560 fee-level countable years (excluding forecasts)

33. To identify regulated £560 fee-level countable years (excluding forecasts) from the individualised file select HESEXCL = 0, HESREG = 1 and HESFEELV = 560. The algorithms for deriving HESEXCL and HESFEELV are given in paragraphs 68-69 and 37 respectively of Appendix 1.

Regulated £0 fee-level countable years (excluding forecasts)

34. To identify regulated £0 fee-level countable years (excluding forecasts) from the individualised file select HESEXCL = 0, HESREG = 1 and HESFEELV = 0. The algorithms for deriving HESEXCL and HESFEELV are given in paragraphs 68-69 and 37 respectively of Appendix 1.

NHS bursaried courses countable years (excluding forecasts)

35. To identify students on NHS bursaried courses from the individualised file select HESEXCL = 0, HESREG = 1 and HESFEELV = NHS. The algorithms for deriving HESEXCL and HESFEELV are given in paragraphs 68-69 and 37 respectively of Appendix 1.

36. We make assumptions about students in receipt of an NHS bursary. Details of these assumptions are given in paragraphs 9 and 10 of Appendix 3.

Foundation degree bridging courses countable years (excluding forecasts)

37. To identify students on foundation degree bridging courses from the individualised file select HESEXCL = 0, HESREG = 1 and HESFEELV = FDBC. The algorithms for deriving HESEXCL and HESFEELV are given in paragraphs 68-69 and 37 respectively of Appendix 1.

Non-regulated fee-level countable years (excluding forecasts)

38. To identify non-regulated fee-level countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESFEELV = OTHER. The algorithms for deriving HESEXCL, HESREG and HESFEELV are given in paragraphs 68-69, 65 and 37 respectively of Appendix 1.

Undergraduate full-time and sandwich medical and dental countable years (excluding forecasts)

39. To identify undergraduate full-time and sandwich medical and dental countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESMED = 1. The algorithms for deriving HESEXCL, HESREG and HESMED are given in paragraphs 68-69, 65 and 59 respectively of Appendix 1.

Undergraduate full-time and sandwich medical and dental countable years (including forecasts)

40. To identify undergraduate full-time and sandwich medical and dental countable years (including forecasts) from the individualised file select HESEXCL = 0 and HESMED = 1. The algorithms for deriving HESEXCL and HESMED are given in paragraphs 68-69 and 59 respectively of Appendix 1.

Undergraduate full-time and sandwich medical and dental assumed countable years

41. To identify undergraduate full-time and sandwich medical and dental assumed countable years from the individualised file select HESCOL4 = 1 and HESMED = 1. The algorithms for deriving HESCOL4 and HESMED are given in paragraphs 80 and 59 respectively of Appendix 1.

Pre-clinical medical countable years (excluding forecasts)

42. To identify medical and dental countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESMED = 1. To obtain the proportion of activity in pre-clinical medicine select REGBODY = 01 and sum the values of PRGA subtracted from 1. The algorithms for deriving HESEXCL, HESREG, HESMED and PRGA are given in paragraphs 68-69, 65, 59 and 54-58 respectively of Appendix 1.

Clinical medical countable years (excluding forecasts)

43. To identify medical and dental countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESMED = 1. To obtain the proportion of activity in clinical medicine select REGBODY = 01 and sum the values for PRGA. The algorithms for deriving HESEXCL, HESREG, HESMED and PRGA are given in paragraphs 68-69, 65, 59 and 54-58 respectively of Appendix 1.

Pre-clinical dental countable years (excluding forecasts)

44. To identify medical and dental countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESMED = 1. To obtain the proportion of activity in pre-clinical dentistry select REGBODY = 02 and sum the values for PRGA subtracted from 1. The algorithms for deriving HESEXCL, HESREG, HESMED and PRGA are given in paragraphs 68-69, 65, 59 and 54-58 respectively of Appendix 1.

Clinical dental countable years (excluding forecasts)

45. To identify medical and dental countable years (excluding forecasts) from the individualised file select HESEXCL = 0 and HESREG = 1 and HESMED = 1. To obtain the proportion of activity in clinical dentistry select REGBODY = 02 and sum the values for PRGA. The algorithms for deriving HESEXCL, HESREG, HESMED and PRGA are given in paragraphs 68-69, 65, 59 and 54-58 respectively of Appendix 1.

New entrants

46. To identify new entrants from the individualised file select HESEXCL = 0 and YEARONE = 1. The algorithms for deriving HESEXCL and YEARONE are given in paragraphs 68-69 and 38 respectively of Appendix 1.